Cognition 231 (2023) 105323

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Doesn't everybody jaywalk? On codified rules that are seldom followed and selectively punished

Jordan Wylie [a,*], Ana Gantman [b,c]

[a] Boston College, USA
[b] Brooklyn College, USA
[c] The Graduate Center, City University of New York, USA

ARTICLE INFO

Keywords:
Morality
Norms
Punishment and blame
Motivation
Third-party punishment

ABSTRACT

Rules are meant to apply equally to all within their jurisdiction. However, some rules are frequently broken without consequence for most. These rules are only occasionally enforced, often at the discretion of a third-party observer. We propose that these rules—whose violations are frequent, and enforcement is rare—constitute a unique subclass of explicitly codified rules, which we call 'phantom rules' (e.g., proscribing jaywalking). Their apparent punishability is ambiguous and particularly susceptible to third-party motives. Across six experiments, ($N = 1440$) we validated the existence of phantom rules and found evidence for their motivated enforcement. First, people played a modified Dictator Game with a novel frequently broken and rarely enforced rule (i.e., a phantom rule). People enforced this rule more often when the "dictator" was selfish (vs. fair) even though the rule only proscribed fractional offers (not selfishness). Then we turned to third person judgments of the U.S. legal system. We found these violations are recognizable to participants as both illegal and commonplace (Experiment 2), differentiable from violations of prototypical laws (Experiments 3) and enforced in a motivated way (Experiments 4a and 4b). Phantom rule violations (but not prototypical legal violations) are seen as more justifiably punished when the rule violator has also violated a social norm (vs. rule violation alone)—unless the motivation to punish has been satiated (Experiment 5). Phantom rules are frequently broken, codified rules. Consequently, their apparent punishability is ambiguous, and their enforcement is particularly susceptible to third party motives.

Humans are extremely concerned with what other humans are doing, especially if they are violating social norms—representations of how most people do and ought to behave (Asch, 1956; Bearden and Etzel, 1982; Goldstein et al., 2008; Miller and Prentice, 1996, 2016; Schultz et al., 2007; Henrich et al., 2003; Henrich & Henrich, 2006; Henrich et al., 2010; FeldmanHall et al., 2014). Those who deviate from norms often experience negative social repercussions (Brauer and Chaurand, 2010; Brauer and Chekroun, 2005; Fehr and Fischbacher, 2003)—sometimes coming directly from the victims themselves, in an attempt to prevent further harm (i.e., second party punishment; Fehr and Gächter, 2000, 2002; Clutton-Brock and Parker, 1995; Crockett et al., 2014). Other times, these costly repercussions come from a third, unaffected party (i.e., third-party punishment; Fehr and Fischbacher, 2004; Buckholtz et al., 2008; Jensen et al., 2007; Riedl et al., 2012) in the form of loss of money (e.g., after an unfair offer in an economic game; e.g., Fehr and Fischbacher, 2003) or via interpersonal consequences like ostracism (Curtis, Robertson, Cockrell, & Fayard, 2021). Humans have also devised a way to punish unwanted behavior via explicit codes of conduct, violations of which incur pre-specified punishments doled out by a third-party institution (e.g., the U.S. legal system; Searle and Willis, 1995). These codes of conduct are designed to proscribe actions (not individuals) and are meant to apply uniformly to all (Weber, 1921/2002; Graeber, 2015; Nadler, 2012; Hannikainen et al., 2021). Yet, they often do not.

Some rules within the system are both frequently broken and rarely enforced (e.g., jaywalking), making it ambiguous whether any given instance of breaking the rule should be or will be punished. As a result, these rules can be invoked at the punisher's discretion, and can be enforced as way to punish norm-violations (i.e., behaviors not proscribed by the system). In the present research, we seek to show that these codified, frequently broken rules represent a unique subclass of rules, which consequently, are enforced in accordance with people's

* Corresponding author at: Department of Psychology & Neuroscience, Boston College, Chestnut Hill, MA 02467, USA.
E-mail address: wyliej@bc.edu (J. Wylie).

https://doi.org/10.1016/j.cognition.2022.105323
Received 8 April 2022; Received in revised form 28 October 2022; Accepted 3 November 2022
Available online 18 November 2022

motivation to punish behaviors that are technically allowed but disliked. Although social and legal rules seem to constitute two separate spheres, we find that people can use codified rules to enforce social ones.

Humans both young and old will incur a cost to punish someone who is acting unfairly or selfishly (Buckholtz and Marois, 2012; Crockett et al., 2014; Hamlin et al., 2011; McAuliffe et al., 2015) to weed out "free riders" who stand to benefit from the collective without sacrificing for it (Fehr and Fischbacher, 2004, Fehr et al., 2002, Raihani and McAuliffe, 2012). This enforces norms of fairness and promotes cooperation, critical for human proliferation and flourishing (Tomasello, 2009, 2014). Although one ostensible purpose of third-party punishment is to promote fairness, there is also substantial evidence that third-party punishment is subject to biases. Features like group membership (Bernhard et al., 2006; Schiller et al., 2014; Yudkin et al., 2016), complexity of population (Marlowe et al., 2008), and physical attractiveness (Li and Zhou, 2014) all have marked influences on to whom and how much punishment is allotted by third parties. People prefer to see strangers, but not close friends and family, punished for severe transgressions (Weidman et al., 2020). This is one hazard of relying on third-party punishment for social regulation. People's alliances and motivations affect who is punished and who is not.

Another way to manage and communicate what behaviors are appropriate in a society is to codify them. Humans forge codes of conduct upheld by institutions, to ensure that people behave as expected and as desired (e.g., legal codes, rulebooks in sports), which can be legitimately punished by the system (rather than interpersonally) when they do not. It is a key feature of rules that people not only know what they are, but that they must apply equally to everyone (e.g., Weber, 1921/2002; Graeber, 2015; Hannikainen et al., 2021). By design, rules are meant to be impersonal, apply universally, and proscribe specific behaviors, rather than condemn individuals or their character (Nadler, 2012; Nadler and McDonnell, 2011). These formalized rules essentially set up an official system for third-party punishment, where instead of relying on representations of what is appropriate (i.e., social norms), proscribed behaviors and the consequences of disregard are explicitly written. This allows for the experience of fairness, general compliance, and expectation of punishment when rules are broken—that is, assuming people perceive the system to be legitimate (Tyler, 2006; Robinson and Darley, 1995; Levi and Sacks, 2009).

Yet we know that in practice, the rules do not apply equally to all. Some people seem to get away with breaking the rules while others do not (e.g., Biernat et al., 1991; Biernat and Fuegen, 2001; Biernat and Manis, 1994; Chawla et al., 2020; Samoylov et al., 2020; Southern Poverty Law Center, 2018; Thompson, 2017). This is particularly evident in the enforcement of rules that are frequently broken and rarely enforced, a kind of rule that we are calling a "phantom rule." Phantom rules are directives that are codified by law (or some other explicit code of conduct), violated frequently, and rarely enforced. Some phantom rules in the U.S. legal system are jaywalking and downloading music without paying for it. These kinds of rules are uniquely consequential (and merit their own name) because judgments of their punishability (i.e., whether it is justifiable to punish someone for violating them) are ambiguous. We call these directives phantom rules because they are mostly *invisible*. We forget these rules exist while everyone is breaking them—unless someone's desire to punish you summons them from obscurity.[1]

As an example, take an umpire in professional tennis. It is their job to uniformly enforce the rules of tennis from a third-party perspective. But, if you know what to look for when watching a professional tennis match, you'll almost always notice coaches giving feedback to the athletes (i.e., coaching) while they compete. Coaching is technically illegal during Grand Slam tennis tournaments, but the rules proscribing coaching are notoriously ignored. Serena Williams is one of the most accomplished tennis players in history; yet, Williams has been repeatedly criticized for her nontraditional tennis attire (see McLaughlin, 2018), and her emotional "tirades", which are technically permissible but disliked by tennis authorities (Clarey, 2009), resulting in a contentious relationship with tennis umpires in general (Raggs and Boren, 2018). In the 2018 U.S. Open, Williams received an unprecedented coaching code violation that ultimately cost her $17,000 (Raggs and Boren, 2018), and possibly the title. Here, we suggest that one reason the umpire may have called the coaching violation was because of a pre-existing motivation to punish Williams for her other norm-violating (but technically legal) behavior.[2] Here, we argue that the umpire invoked coaching, which is established within the codified rules of professional tennis, as a way to punish Williams for other transgressions not technically proscribed by the rules. In this way, phantom rule enforcement tells us something new about third-party punishment: We sometimes use codified rules to punish people for transgressions *beyond* the purview of those rules.

We argue that it is especially ambiguous whether a phantom rule should be enforced at any given instance of rule breaking. This ambiguity in enforcement makes these rules particularly susceptible to the same kinds of biases that occur in third-party punishment of selfish (though technically legal) behavior (e.g., Bernhard et al., 2006; Schiller et al., 2014; Yudkin et al., 2016) and are evident in the massive racial inequality in the American criminal justice system (Arnold et al., 2018; Bender et al., 2021; Huebner and Bynum, 2008; Pierson et al., 2020). But, because these are codified rules, they also provide a seemingly legitimate reason to have the system that put the rule in place be responsible for administering the punishment. To return to our tennis example, it might be difficult to punish Williams for her norm-violations on court, but there is a recognized avenue for punishment for coaching. In this way, the rules that were designed to apply equally to all, can be invoked to fulfill punishment motives activated by other behaviors (e.g., norm violations), which are otherwise more difficult and costly to punish.

1. Present research

In the present research, we investigated the existence and consequences of frequently broken, rarely enforced rules, which we call phantom rules. First, we sought behavioral evidence of motivated phantom rule enforcement in a tightly controlled setting (Experiment 1). Then we used experimental surveys to show that motivated judgments of phantom rules generalize to multiple legal rules, are distinct from more prototypical legal rules (and social norms in a supplementary experiment) and are particularly susceptible to motivated enforcement (Experiments 2–5). Together, the present research identifies the existence and significance of a subset of explicit rules that are rarely followed and are both enforced and judged in a motivated manner. All experiments except for Experiment 2 were pre-registered; data, materials, code for analyses, for all experiments, are available on the Open Science Framework.[3]

2. Experiment 1

Our first aim was to find behavioral evidence of motivated punishment for phantom rule violations in a highly controlled economic game.

[1] According to Casper the Friendly Ghost universe, ghosts appear when they have unfinished business. Phantom rules are sometimes enforced when one person has unfinished business with another, invoked to try to finish up that business.

[2] While many factors may have contributed to the coaching penalty, including the strictness of the umpire himself, we see this event within the broader lens of Williams' consistently facing discrimination in professional tennis. In a way, this is the point. Phantom rules give a veneer of legitimacy to someone looking to exact punishment.

[3] https://osf.io/dhmjx/

To do this, we employed a modified version of the Dictator Game (Forsythe et al., 1994), which we refer to here and in the experiment as The Sharing Game. In the Sharing Game, we pit rational, economically beneficial decision-making against the desire to punish a selfish (vs. fair) "dictator" similar to other work examining inequality aversion and third-party punishment (e.g., Balafoutas and Nikiforakis, 2012; Fehr et al., 2002; Fehr and Fischbacher, 2002, 2004; FeldmanHall et al., 2014; Goette et al., 2006; Raihani and McAuliffe, 2012). Here, the only possible avenue for punishment was via the enforcement of the predefined rules of the Sharing Game. Players were told about these rules and their punishments rates. Critically, one rule was frequently broken and rarely enforced (i.e., a phantom rule). We hypothesized that phantom rules would be enforced (which forfeits economic benefit) to a greater extent when people were motivated to punish the "dictator" because of behavior outside of the purview of the rules (i.e., the "dictator" was selfish rather than fair, which is technically allowed though known to evoke the desire to punish; FeldmanHall et al., 2014).

2.1. Methods

2.1.1. Participants

We recruited a total of 526 individuals from Prolific. We preregistered exclusion criteria to remove bots, individuals who failed game rule comprehension questions, individuals who did not believe the game was played with other real participants, and individuals who self-reported not paying attention during the game. This left us with a final sample of 409 (M_{age} = 33.17, SD_{age} = 9.80, 183 females, 214 males, 10 other, 2 did not report) consenting participants. We pre-registered data collection of 425 participants (and collected about ~20% additional to accommodate messy online data and exclusions), which was the amount that we had funds to pay. We also conducted a power analysis to determine whether our sample size was sufficient to detect a 10% difference between the two conditions. We used G*Power (Faul et al., 2009) and specified a two-tailed test with an odds ratio of 1.49, alpha set to 0.05, and power set to 80%. A sample size of 213 would be sufficient for 80% to detect a 10% difference between conditions. The only inclusion criterion was current United States residence. After completion of the study, participants were compensated $2.25 for their time, and each participant who completed the game received a $0.75 bonus.

2.2. Materials

2.2.1. Sharing game

We designed a modified three-player dictator game (Forsythe et al., 1994), and called it the Sharing Game, with an "Allocator" who decides how much of the given 5 points (1 point translates to $0.05) to split between themselves and the Receiver. An Observer decides whether the entire game was valid or invalid (akin to an umpire or referee). Both the Allocator and the Receiver are pre-programmed into the game set-up, making the Observer the only real participant playing the game. After an explanation of each of the three possible roles (for sample instruction page see Fig. 1A), participants were assigned to the Observer role. The

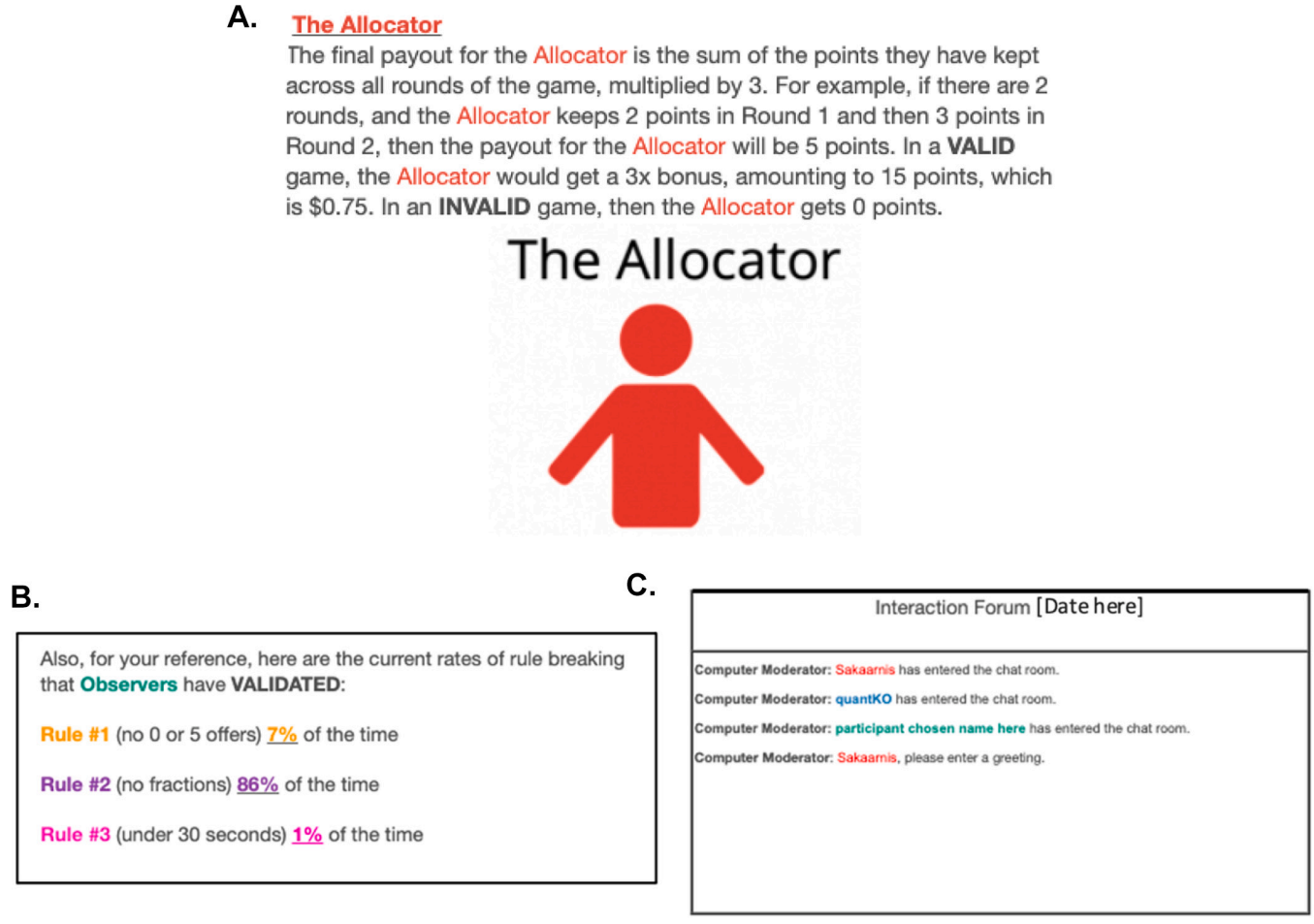

Fig. 1. Fig. 1A is an example of the instruction screens for each of the roles. Fig. 1B is a screen capture of the rule and norm manipulation shown to each participant. Fig. 1C is a screen capture of the chat forum that we adapted from Yudkin et al., 2016. The red and blue screen names were chosen at random and were the same across participants. The green represents where participants would see their own screen name that they selected. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

Observer was asked to decide whether to deem a game valid, which would result in payout of the offers with a 3 times multiplier, or invalid, which would result in loss of points for the Allocator and no multiplier for the Observer or Receiver. Observers were instructed to base their validation decision on whether the actions of the Allocator followed the explicitly stated rules of the game. Participants were told that rule enforcement was up to their discretion.

The explicit rules were: (1) no 0 or 5 offers, (2) make no fractional offers (e.g., 2.5), and (3) make offers within 30 s (see Fig. 1B). Each rule also had information about how often those rules were enforced by previous observers. The Observer was asked to make their single game validation decision after watching a series of five offers made by the Allocator. To test whether external motivations influence phantom rule enforcement, we manipulated selfishness of the five offers from the Allocator. Participants were randomly assigned to view five sequential rounds of a selfish Allocator, who kept an average of 3.9 points, or a fair Allocator, who kept an average of 1.3 points of the total 5 points per round available to split. Both the fair and the selfish Allocator break the same rarely enforced rule of offering a fraction of a point on the third round of the game. The game rules, enforcement rates, and type of rule violation were identical across conditions.

We also had participants run through the rules of the game one additional time through a simulated chat adapted from Yudkin and colleagues (see Fig. 1C; Yudkin et al., 2016). This allowed us to introduce the mock characters. We used bogus chat names for the Allocator ("Sakaarnis") and Receiver ("quantKO"). The Observer entered their own chat name, and all "players" wrote a greeting. Each of the rules were explained with images (see Fig. 2) and comprehension of the rules of the game was assessed with three multiple choice questions. Participants had to get these questions right to proceed to the game. Full game instructions are available on OSF.

2.2.2. Comprehension check

Given the complexity of the game, we asked participant three comprehension questions to gauge understanding of the game instructions and the reward payout for validated games. Each participant was asked to indicate how many points the Allocator would receive if a game was invalidated, and how many points the Observer would get if the game was invalidated or validated. Participant had three chances to answer each question correctly. People who required more than one try to answer any comprehension question were not included in analyses (as pre-registered).

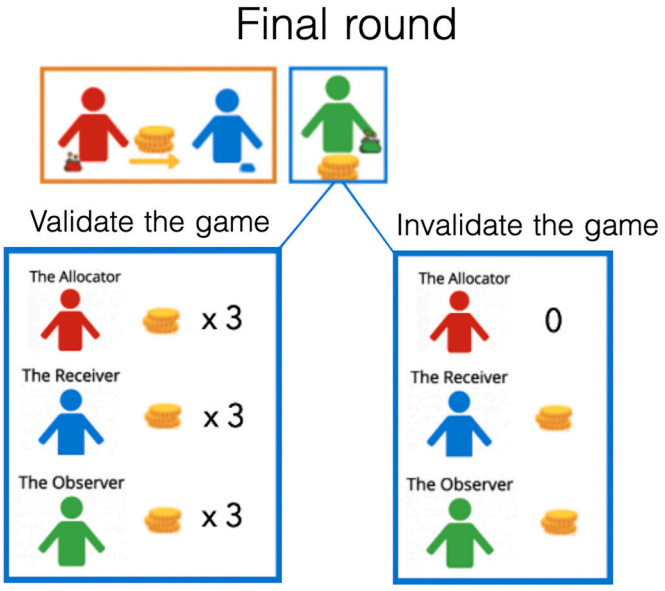

Fig. 2. After the observing a series of offers, the fifth and final trial required the Observer to make a validation decision. If the game was deemed valid, Observers were told that all players receive a 3× payout. If the game was deemed invalid, the Allocator received no payment and the Observer and Receiver receive no multiplier. Thus, it was in the interest of the Observer to ignore rules (regardless of their descriptive enforcement norms) and deem games valid.

2.2.3. Manipulation check

To assess whether the manipulation of average offers affected judgments of the Allocator, we asked: "What was your impression of Sakaarnis (The Allocator)?" and "What was your impression of quantKO (The Receiver)?" rated from 1 = *Extremely Negative* to 7 = *Extremely Positive*.

2.2.4. Validation judgment

Observers watched and progressed the game forward for four trials. They were then alerted of the fifth and final trial where they made their final validation decision for the full game. For the fair condition, the final decision read: "Sakaarnis (The Allocator) kept 2 points and gave quantKO (The Receiver) 3 points." Participants were shown a "validate" and an "invalidate" option that had final payout information available (as a reminder that choosing "validate" would end in a 3× payout). For the selfish condition, the final decision read: "Sakaarnis (The Allocator) kept 3 points and gave quantKO (The Receiver) 3 points." The same "validate" and "invalidate" options were presented, but the payout information differed given the different amounts allocated between the trials.

2.2.5. Rule stickler judgment

We also included a single item from the Right-wing Authoritarianism scale (Altemeyer, 1998) to assess whether some individuals were "sticklers" for the rules. The selected item was "Obedience and respect for authority are the most important values children should learn.", rated from 1 = *strongly disagree* to 7 = strongly agree.

2.3. Procedure

Following informed consent, participants were randomly assigned to a condition with either a selfish or fair Allocator and given the three rules to enforce at their discretion (see Fig. 2). Participants were tested on their comprehension of the rules of the game and their role prior to beginning the game offers. Critically, we told participants that offers with a decimal (e.g., 1.5 points) were against the rules, but also that previous participants had deemed games with this rule violation valid anyway. All participants saw an Allocator (either a selfish or a fair) who broke this rule. Observers were required to acknowledge that the offer was made prior to moving to the next offer to maintain attention during the full task. After the fifth and final offer, Observers decided on whether to deem the game valid or invalid, followed by an explanation for their decision. Participants then made a global liking rating of the other "players" and answered questions about whether they believed they were playing with real players, whether they paid attention, and answered demographic questions. All participants were debriefed, compensated for their time, and received the additional payout (regardless of whether they deemed the game valid or invalid) after completion of the experiment.

2.4. Results

The primary analyses for Experiment 1 used linear regression models. All analyses were conducted using R statistical analysis software (R Core Team, 2019). For the target validation decision, we used logistic regression to accommodate the binary nature of the response. For manipulation check items, we use linear regression. Given the structure of these data, no tests were run with random effects included.

2.4.1. Manipulation check

After the final trial, participants rated their impression of the two other players. Results suggested that the manipulation was successful. Allocators in the selfish condition were seen as less likeable than those in the fair condition, $b = -1.67$, $SE = 0.16$, $t(407) = -10.16$, $p < .001$, $r = -0.45$, *95% CI* [−1.99, −1.34], $OR = 0.19$. Condition did not significantly affect the ratings of the Receiver, $b = -0.10$, $SE = 0.12$, $t(407) = -0.83$, $p = .41$, $r = -0.04$, *95% CI* [−0.33, 0.13], $OR = 0.91$.

2.5. Validation decision

Participants were more likely to invalidate the game, invoking the phantom rule of fractional offers and forfeiting their potential extra payout, when Allocators were selfish (vs. generous), $b = -0.55$, $SE = 0.20$, $z = -2.72$, $p = .007$, $r = -0.15$, *95% CI* [−0.92, −0.13], $OR = 0.58$; our desire to punish behavior outside of the purview of the rules partly determines whether punishment for the very same rule violation is doled out or not (see Fig. 3). We also tested (as pre-registered) a model that included condition, rule "stickler", and their interaction term. However, no significant effects emerged, $ps > 0.13$.

2.6. Discussion

We found that participants were more likely to enforce a phantom rule to punish the Allocator in the Sharing game when the Allocator was selfish rather than fair (which is technically allowed). We created a phantom rule by telling all participants that some of the rules were infrequently enforced, and we chose a rule that we thought did not convey much social meaning to break. Notably, participants did not uniformly enforce the rule. Instead, participants were more likely to enforce the rule, at a cost to themselves, when the Allocator acted selfishly (vs. fairly). In other words, the Allocator elicited a punishment motivation by acting selfishly (Falk et al., 2003, 2008; Fehr and Fischbacher, 2002)—a behavior that is technically allowed, but outside the purview of the rules. Unlike we predicted, our single item index of being a rule stickler did not predict validation decisions. Overall, in a highly controlled economic game, we found evidence for the motivated enforcement of rarely enforced (i.e., phantom) rules. Importantly, however, we do not yet know how unique this might be to phantom rules and whether it applies more broadly outside of a controlled economic game.

3. Experiment 2

We looked to the legal domain to test whether our conception of phantom rules is recognizable to others, and whether people readily distinguish phantom rule violations from other violations. In particular, we sought to contrast phantom rule violations with non-codified social norm violations because they differ on the two defining features of phantom rules, namely legality and frequency of violation. Social norms are representations of the most common or desirable thoughts and behaviors of a group (Miller and Prentice, 1996, 2016), which people tend to use as a guide for their own behavior (Cialdini and Goldstein, 2004; Miller and Prentice, 1996, 2016). Phantom rules are legally codified and frequently broken, whereas social norms are not legally codified, and frequently adhered to.

For the purposes of this experiment, we adopted a narrow definition of social norms (see Bicchieri et al., 2011 for detailed discussion)—one where the violation of such a norm is socially consequential, but no edict exists to forbid it. Unlike phantom rules, social norms exert meaningful influence on individual behavior (Asch, 1956; Bearden and Etzel, 1982; Goldstein et al., 2008; Miller and Prentice, 1996, 2016; Schultz et al., 2007), at least partly because people who deviate from norms experience negative social repercussions (Brauer and Chekroun, 2005; Brauer and Chaurand, 2010). That is, unlike violations of phantom rules, social norm violations, which are not technically against any codified rule, tend to be interpersonally consequential. They are also enforced ad hoc via social means, such as ostracism, "canceling", and call-out culture (e.g., boycotting an individual's work or broadcasting your dislike on social media). In this way, phantom rules are the 'inverse' of social norms. Phantom rule violations are typically socially inconsequential but technically illegal, and thereby punishable by the system that put the rule in place.

In Experiment 2, we sought to identify phantom rules in the U.S. legal system, by asking people to categorize behaviors as either legal or illegal, and common or uncommon. In an exploratory manner, we also examined the extent to which people find those same behaviors to be morally acceptable, enforced (i.e., is this a rule whose violation brings about consequences), and deserving of blame (i.e., is this a rule whose violation deserves blame).[4] Experiment 2 allowed us both to identify whether people recognize phantom rules (vis-à-vis social norms), and to select a set of phantom rules for further study. We hypothesized that people would be able to more frequently categorize phantom rule violations as illegal (vs. legal) and frequent (vs. infrequent) compared to social norm violations. Then, we used these criteria to identify the most prototypical phantom rules in the U.S. legal system for further investigation in Experiments 3–5.

3.1. Methods

3.1.1. Participants

We recruited a total of 123 individuals ($M_{age} = 32.00$, $SD_{age} = 9.14$, 42 females, 69 males, 7 other, 5 not reported) from Amazon's Mechanical Turk. Due to the task demands, we noticed that many failed to complete the survey and so we collected 15 additional participants prior to analyzing the data. We only analyzed the data from the individuals who completed the entire task (all six blocks), which left us with a final sample of 96 ($M_{age} = 32.54$, $SD_{age} = 9.08$, 34 females, 57 males, 5 other) consenting participants. While this experiment was exploratory, we conducted a power analysis using G*Power (Faul et al., 2009) which indicated that a sample of 108 participants gave us 80% power to detect a medium effect ($w = 0.30$). The only inclusion criterion was current United States residence to ensure familiarity with the norms and laws listed in the experiment. After completion of the study, participants were compensated $1.25 for their time.

3.2. Materials

3.2.1. Stimuli

We generated a list of phantom rules and a list of social norms. The first list contained a number of rules that we thought were good face-valid examples of phantom rule violations (e.g., "jaywalking", "using marijuana"). The second list contained social norm violations that ranged in commonality, extremity, and valence. Each of the 25 rule violations (11 social norms, and 14 phantom rules) was a short phrase ranging from one to eight words in length. See supplementary materials for the full list.

3.2.2. Sorting task

Participants saw a single list of all rule violations (i.e., both phantom rule and social norm violations) in random order. They were asked to drag and drop each rule into one of two possible dimensions within a category. For example, for legality, participants were asked to sort each rule into either the "legal" bin or the "illegal" bin. Each participant saw

[4] Pilot data ($N = 260$) reported in the Supplemental Material that was collected as part of a later experiment suggests that unlike phantom rule violations, legal rule breaking of strict laws is low in frequency ($M_{\text{number of people out of 100}} = 22\%$), near ceiling for legal status ($M = 6.66$), and high for moral wrongness ($M = 6.00$). All scales had a maximum rating of 7. Full descriptive statistics reported in the supplemental material.

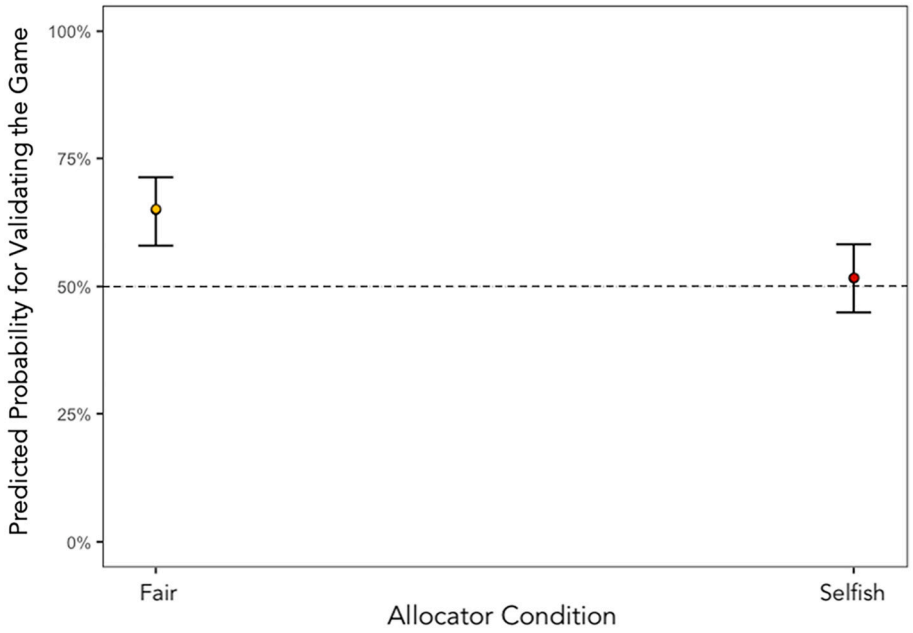

Fig. 3. Predicted probability of selecting validate game (vs. invalidate) for the Fair and Selfish Conditions. Notably, the error bars for the Selfish condition pass through the 50% mark (the gray dashed line). Error bars represent 95% Confidence intervals.

five total binary categories: legality ("illegal" vs. "legal"); frequency ("common" vs. "uncommon"); enforcement ("experience negative consequences" vs. "experience NO negative consequences"); wrongness ("deserves punishment" vs. "does not deserve punishment"); morality ("morally acceptable" vs. "moral unacceptable"); and a category of their own labeling (see Fig. 4). Participants sorted one dimension at a time (dimensions presented in random order), and all rule violations were sorted before participants could move forward.

3.3. Procedure

Following informed consent, participants sorted the list of 14 phantom rule and 11 social norm violations in the sorting task. After completing the entire task, participants were then asked to answer general demographic questions concerning their age, gender identity, and race. All participants were paid for their participation and were debriefed at the end of the study.

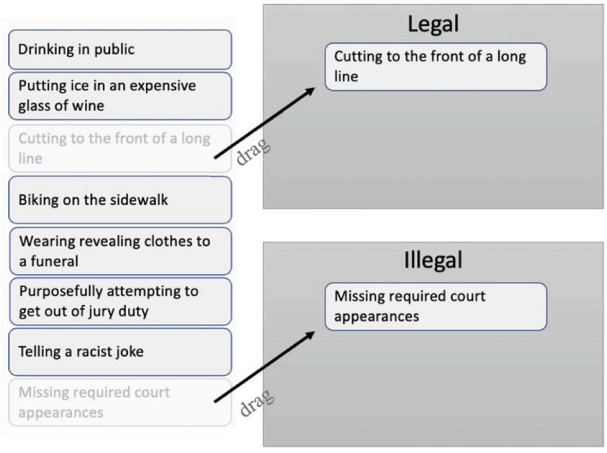

Fig. 4. Example of the sorting task procedure for the legality category. Participants sorted the lists of randomly ordered rules by dragging and dropping a rule into one of the two bins, one rule at a time. Here we depict a social norm violation being sorted into the "legal" bin and a phantom rule violation being sorted into the "illegal" bin.

3.4. Results

The primary analyses for Experiment 2 utilized generalized linear mixed-effect models given the binary nature of the dependent variables.[5] All analyses were conducted using R statistical analysis software (R Core Team, 2019), using the 'glme4' (Bates et al., 2015) and the 'lmerTest' packages (Kuznetsova et al., 2017) for model *p*-values using Satterthwaite's approximation. For each dependent variable (legal, frequent, moral, enforce, deserve, and self-labeled), we entered random intercepts for participants and specified rule category (phantom rule vs. social norm violation) as the predictor.[6]

3.5. Main analyses

We first investigated whether phantom rule violations were more often sorted into the illegal compared to legal bin as per our definition of phantom rules. Results supported this prediction, people more often selected the "illegal" option for phantom rule violations compared to social norm violations, $b = -0.76$, $SE = 0.04$, $z = -16.60$, $p < .001$, $r = -0.21$, *95% CI* [−0.85, −0.69], $OR = 0.47$.[7] Next, we examined if the full list of phantom rule violations were perceived as more common compared to the full list of social norm violations. Again, results supported this prediction, phantom rule violations were more frequently categorized as commonly occurring than social norm violations, $b = 0.15$, $SE = 0.04$, $z = 3.46$, $p < .001$, $r = 0.04$, *95% CI* [0.07, 0.24], $OR = 1.17$. Finally, for the exploratory dimensions, results suggested that phantom rule violations were less frequently categorized as morally acceptable ($b = -0.12$, $SE = 0.04$, $z = -2.86$, $p = .004$, $r = -0.03$, *95% CI* [−0.21, −0.04], $OR = 0.88$), more commonly enforced ($b = 0.11$, *SE*

[5] For all analyses Phantom Rule Violations are coded as 1 and Social Norm Violations (or Prototypical Legal Violations) are coded as −1.

[6] Random effects for stimuli did not meet the clustering cutoff (ICC > 0.10) used across experiments and were not included.

[7] The pattern of results from the full sample ($N = 119$) followed the same general pattern as the results for only the individuals who completed the experiment.

= 0.04, $z = 2.50$, $p = .012$, $r = 0.03$, *95% CI* [0.02, 0.20], $OR = 1.11$), and more deserving of punishment ($b = 0.19$, $SE = 0.04$, $z = 4.51$, $p < .001$, $r = 0.05$, *95% CI* [0.11, 0.28], $OR = 1.21$) likely owing to their codified status. Altogether, these results suggest that people are able to recognize the key differences between phantom rule and social norm violations, that phantom rules are both legally codified and frequently broken.

3.5.1. Rule selection

3.5.1.1. Selection process. It was also the aim of Experiment 2 to select a subset of phantom rule and social norm violations for further study. We selected six phantom rule and six social norm violations with relatively high consensus on the two definitional categories of phantom rules: legality and frequency of violation. Thus, we selected six phantom rule violations that were consistently recognized as illegal and frequent, and six social norm violations that were consistently recognized as legal and infrequent. That is, of the participants who completed the whole survey ($N = 96$), 74.5% identified these six phantom rule violations as *frequent* compared to 59.7% for the six chosen social norms. For the legality dimension, 58.7% of the final sample correctly identified these six phantom rules as legal compared to 23.7% for the six social norms selected (see Table 1[8]). We selected six phantom rule violations we felt confident that future participants could recognize as illegal but frequent, and six social norm violations that we felt confident future participants could recognize as legal but infrequent. Additionally, in doing this, the dimension of morality switched direction: Phantom rule violations were more likely to be categorized as morally acceptable (55%) compared to social norm violations (46%).

3.5.1.2. Subset validation. To ensure that the selected rules were evidencing the same predicted patterns of results, we re-ran the models first with our definitional categories of legality and frequency, and then with our dependent variable of interest, morality. Our subset of phantom rule violations were more frequently sorted as common ($b = 0.30$, $SE = 0.07$, $z = 4.28$, $p < .001$, *95% CI* [0.16, 0.44], $OR = 1.35$) and as illegal ($b = -0.83$, $SE = 0.07$, $z = -11.96$, $p < .001$, *95% CI* [−0.96, −0.70], $OR = 0.44$) than our subset of social norm violations (see Table 2).

3.6. Discussion

We found that participants recognized that phantom rule violations are more illegal and more frequent than social norm violations. We also selected and validated a subset of rules for future study based on which behaviors had the greatest consensus on these two definitional criteria. We selected the phantom rule violations that most participants recognized as legal rules that are frequently violated. We selected the social norms that most participants recognized to be outside the legal domain, but rarely violated. Within our selected subset of rules, phantom rule violations were judged to be more morally acceptable than social norm violations. It is worth noting that when we restrict to the subset of phantom rule and social norm violations that best fit our criteria for identifying phantom rules, the phantom rule violations are more likely to be sorted as morally acceptable than the social norm violations. When we remove behaviors that don't meet these criteria in the minds of participants, we see a change in moral acceptability. We suspect that this happened for a few reasons. People may have to *know* that a behavior is both illegal and descriptively normative for it to lack prescriptive force. In addition, multiple factors determine participants judgments of the moral acceptability of these actions, some of them in competing directions. For example, moral and legal judgments are closely tied (Dunlea and Heiphetz, 2021; Flanagan & Hannikainen, 2022), which would suggest a positive relationship between what actions are illegal and what actions people find to be wrong. What is descriptively normative and what is morally acceptable are also closely related (Eriksson et al., 2015; Lindström et al., 2018), which would suggest a positive relationship between what is common and what is acceptable. Rule violations that people know to be frequent tend to also be seen as more morally acceptable, supporting the idea that "what is common is acceptable" (Lindström et al., 2018). And further, the descriptive normativity of these rules suggest that phantom rules likely do not meet the criteria of being moral norms at all (see Bicchieri, 2006; Brennan et al., 2013). These rules are not unconditionally followed, and therefore their violations are not likely subject to moral blame under non-motivated conditions. Nonetheless, moral acceptability was not a factor in our selection of the final set of phantom rules for further study.

It is also part of our definition of phantom rules that they are rarely enforced. We asked about enforcement in a broad sense and did not find differences in perceived rates of enforcement. In a pilot experiment ($N = 161$; see Experiment S1 in the Supplemental Material), we did find that phantom rules differ in the kinds of punishments that are appropriate following their violation (legal punishment for breaking the law, social punishment for breaking a social norm). In Experiment 3, we compared phantom rules to other codified rules—specifically, more prototypical legal violations to better understand how they differ. Phantom rule violations occupy a puzzling normative space; it is explicitly forbidden to break these rules but also descriptively normative to do so—unlike more prototypical legal rules.

4. Experiment 3

Experiment 3 sought to investigate the moral and punishment implications of phantom rule violations compared to more prototypical laws in the United States. Critically, by nature of their codified status, phantom rule violations are subject to the same penalties as other more prototypical rules and laws. However, unlike other laws governing similar behaviors, phantom rules are also commonplace to break and judged to be relatively morally inconsequential, potentially resulting in a lack of legitimacy. Rule legitimacy, which is the appropriateness of rule or law enforcement, reflects the belief that authorities have the right to enforce those rules or laws (Tyler, 2006). Given evidence that phantom rule violations are both frequently broken and not particularly morally consequential, we reasoned that, despite their legal status, phantom rule violations would lack the same prescriptive force as prototypical laws. Altogether, we predicted that people would judge phantom rule violations as less morally wrong, deserving of blame and punishment, and less legitimate than more prototypical laws.

We also sought to test a boundary condition on the recognizability of phantom rules. While we think that some people view phantom rule violations as morally acceptable, this can only be true for people who see some wiggle-room in terms of which rules must be followed in the first place. As such, people with a general proclivity to view rules as important to follow *de se* may not find phantom rule violations to be appreciably different from other kinds of violations. Accordingly, we also investigated whether two conceptually related individual differences influenced judgments of phantom rules: tight cultures (TLC) and right-wing authoritarianism (RWA). These individual differences tap into an individual's tendency to prefer more conventional, traditional, and regulated behavior (Gelfand et al., 2011; Jost et al., 2003; Wilson, 1973).

The tight-loose culture measure captures individual perceptions of how stringently their culture is organized (Gelfand et al., 2006; Gelfand et al., 2011). That is, loose cultures exhibit high tolerance for counter normative behavior, while tight cultures are intolerant of deviant social behavior and have strict social norms (Gelfand et al., 2011). Right-wing authoritarianism is a personality dimension that is characterized by individuals organizing their world along ingroup/outgroup lines and a

[8] *t*-tests were conducted on the percentage scores for legality and frequency of violation for phantom rule vs. social norm violations, both were significant, *p*'s < 0.01. These summary data are found on the project's OSF page.

Table 1
Breakdown of the percentages for violations selected for future experiments.

	Legality (yes, illegal)	Frequency (yes, common)	Rule Enforcement (yes, enforced)	Deservingness (yes, deserving)	Moral Acceptance (yes, acceptable)
Phantom Rules	58.7%	74.5%	56.9%	51.2%	55.0%
Social Norms	23.7%	59.7%	60.4%	53.4%	46.0%

Table 2
Selected phantom rule and social norm violations.

Phantom Rule Violations	Social Norm Violations
Downloading music off of any place where you do not pay to	Yelling in a meeting with your boss
Jaywalking	Ignoring a dress code at a bar or restaurant
Biking on the sidewalk	Telling a racist joke
Viewing an R-rated movie before age 13	Talking during a movie
Using marijuana	Cutting to the front of a long line
Leaving your car's brights on when another car is within view	Walking up the wrong side of the stairs during commuter rush hour

predilection for the maintenance of traditional structure (Adorno et al., 1950; Altemeyer, 1981, 1988, 1998; Whitley Jr, 1999). Individuals who are high in these dispositions share an inclination to see rules as necessarily important.

We reasoned that those high in TLC and RWA would be more sensitive to what people are *supposed to do* (i.e., prescriptive norms) than what we people *actually* do (i.e., descriptive norms). We hypothesized that only those low in TLC and RWA would be sensitive to differences between phantom rule and social norm violations. This boundary condition is important for understanding the implications of phantom rules for morality: Phantom rule violations are only morally acceptable if individuals perceive that at least *some* rule violations are acceptable more generally. In Experiment 3, we directly compared phantom rule to prototypical legal violations to investigate whether phantom rule violations elicit distinct moral, blame, and punishment response patterns from prototypical legal violations, and examined whether this was only true for individuals low in TLC and RWA.

4.1. Method

4.1.1. Participants

We recruited 167 individuals from Prolific (80 females, 82 males, 3 other, 2 not reported, 79.00% White, 16% People of Color, 5% Other or did not report, $M_{age} = 34.6$, $SD_{age} = 11.41$), and compensated them $2.71 for their time. We aimed to analyze data from at least 145 participants to mirror the analyses and power analysis conducted in the pilot Experiment S1 in the supplemental material. As such, our recruitment goal was set at 165 participants in order to account for a ~ 15% attention check failure rate. Following our pre-registered data quality plan, we excluded any participant who failed any of our attention checks, leaving a total sample size of $N = 155$ (74 females, 76 males, 5 other, 78% White, 16.8% People of Color, 5.2% Other or did not report) for all subsequent analyses.

4.1.2. Design

To explore the extent to which phantom rule and prototypical legal violations elicit different patterns of judgments, we used a fully within-subjects design with one two-level factor (Rule Type: Phantom Rule vs. Prototypical Law). Participants read fourteen total vignettes in randomized order with no repeating scenarios. Seven contained phantom rule violations, and seven contained social norm violations.

4.2. Materials

4.2.1. Stimuli

Participants read 14 short vignettes (7 phantom rule and 7 prototypical violations; see Table 3). We pre-registered and used the same phantom rule violations as used in Experiment S1 (see Supplemental Material) but with one additional phantom rule (proscribing loitering). We used 7 prototypical laws from a pilot experiment (see Experiment S2 Supplemental Material), which we selected to be similar in content to phantom rules but descriptively normative to follow (i.e., infrequently violated). Prototypical laws included: cocaine use, running a red light, stealing, impersonating another person, distributing others' private content on the internet, exposing oneself in public, and drunk driving.

4.2.2. Phantom rule validation

We wanted to ensure that our new sample and the addition of the loitering rule also evidenced differences between phantom rule and prototypical legal violations on the definitional aspects of phantom rules. We asked: "How illegal is it to [phantom rule/prototypical legal violation] in the US?", rated from 1 = *Not at all illegal* to 7 = *Absolutely illegal*; and a frequency item: "How many people out of 100 would [phantom rule/prototypical legal violation]?"

4.2.3. Moral, blame & punishment judgments

The four target judgments included one blame judgment ("How deserving of blame is someone who [phantom rule/prototypical legal violation]?") item rated from 1 = *Not at all* to 7 = *Very much*, one moral character judgment ("How morally bad is a person who [phantom rule/prototypical legal violation]?") item rated from 1 = *Not at all* to 7 = *Extremely*, one legal punishment[9] ("How justified is it for legal system to be involved in this situation?") rated from 1 = *Not at all* to7 = *Completely*, and one measure of moral action judgments ("How morally bad is it to [rule violation]?"), rated on a scale of 1 = *Not at all* to 7 = *Extremely*.

4.2.4. Legitimacy judgment

We included a single face-valid item to measure legitimacy ("How legitimate is the law against [rule violation]?"), rated on a scale of 1 = *Not at all* to 7 = *Extremely* to measure legitimacy judgments.

4.3. Moderators

4.3.1. Tight loose culture (TLC) scale

We also included a measure of cultural tightness or looseness to assess the degree to which individuals view the society they live in as a tight culture. Representative items include "In this country, if someone acts in an inappropriate way, others will strongly disapprove", and "People agree upon what behaviors are appropriate versus inappropriate in most situations this country", rated from 1 = *strongly disagree* to 7 = *strongly agree* (Gelfand et al., 2011). Based on the internal reliability, a single TLC variable was created by collapsing across the six items (Cronbach's $\alpha = 0.68$).

4.3.2. Right-wing authoritarianism (RWA) scale

Ten items were selected from the right-wing authoritarianism scale (Altemeyer, 1998; Rattazzi et al., 2007; Zakrisson, 2005) based on

[9] Participants may have interpreted this item as a legal entity intruding the freedom of the individual. To ensure that tested the correct interpretation of this question, we also included a dichotomous police interaction variable in subsequent experiments.

Table 3
Phantom rule and prototypical legal violations.

Phantom Rule Violations	Prototypical Legal Violations
Downloading music off of any place where you do not pay to	Posting private images online
Jaywalking	Running a red light
Loitering	Exposing oneself in public
Using marijuana	Using cocaine
Leaving your car's brights on when another car is within view	Driving while drinking alcohol
Biking on the sidewalk	Impersonating another person

theoretical relatedness (authoritarian aggression and submission items; Rattazzi et al., 2007). This shortened scale measures the extent to which individuals view authorities and social norms as legitimate with representative items such as, "Obedience and respect for authority are the most important values children should learn," rated from 1 = *strongly disagree* to 7 = strongly agree. Again, a single RWA variable was created by collapsing across the ten items (Cronbach's α = 0.97).

4.3.3. Demographics

Lastly, we asked participants to indicate their gender, age, and political orientation.

4.4. Procedure

Consenting participants completed the experiment via survey, presented online using Qualtrics survey software (Qualtrics, 2018). Participants first saw either a phantom rule or prototypical legal violation and were then asked to provide ratings on a seven-point Likert scale. After completing each of the judgments for each of the randomly presented fourteen scenarios, participants then completed the Right-Wing Authoritarianism Short Scale and the Tight-Loose Culture Scale followed by demographic questions. Participants were then debriefed.

4.5. Results

The analysis plan followed our pre-registered plan and the strategy used in a supplementary study (Experiment S1 in the Supplemental Material): all analyses utilized linear mixed models that specify random intercepts for participants. For this experiment, we excluded stimuli as a random effect because the ICC for that random effect was less than the pre-registered cut-off of 0.10 for critical models.[10] As pre-registered, we first examined whether the dependent variables could be collapsed into indices. Theoretically related variables measuring punishment, blameworthiness, and the two morality judgments reliably assessed the same construct (Cronbach's α = 0.94). Thus, we collapsed those items into a moral relevance index. Correlations between the dependent variables and individual difference measures are included in the supplementary material (see Tables S10-S11 in the Supplemental Material).

4.6. Main analyses

4.6.1. Phantom rule validation

We examined whether participants would rate phantom rule violations as *less* illegal but *more* frequent compared to more prototypical laws. We found strong evidence for both hypotheses. Phantom rule violations ($M_{phantom\ rule}$ = 4.71) were seen as less *illegal* compared to prototypical legal violations ($M_{prototypical\ rule}$ = 6.30), $b = -0.80$, $SE = 0.03$, $t(2007) = -26.39$, $p < .001$, $r = 0.51$, *95% CI* [−0.85, −0.74]. Additionally, their violations ($M_{phantom\ rule}$ = 44) were seen as more *frequent* compared to prototypical legal violations ($M_{prototypical\ rule}$ = 22), $b = 11.01$, $SE = 0.44$, $t(2007) = 24.86$, $p < .001$, $r = 0.49$, *95% CI* [10.14, 11.88].

4.6.2. Collapsed morality, blame, & punishment

Next, we examined the extent to which phantom rules and prototypical laws differ for judgments of moral relevance (see Fig. 5). We predicted and found that phantom rule violations ($M_{phantom\ rule}$ = 3.61) were less morally relevant than prototypical legal violations ($M_{prototypical\ rule}$ = 5.84), $b = -1.11$, $SE = 0.03$, $t(2006) = -39.78$, $p < .001$, $r = 0.66$, *95% CI* [−1.16, −1.06].

4.6.3. Legitimacy

We then tested whether phantom rules and prototypical laws differed in how legitimate to enforce they seem (see Fig. 5). Again, we predicted and found that phantom rules ($M_{phantom\ rule}$ = 4.03) are viewed as less legitimate to enforce than prototypical laws ($M_{prototypical\ rule}$ = 5.99), $b = -0.98$, $SE = 0.03$, $t(2006) = -30.60$, $p < .001$, $r = 0.56$, 95% *CI* [−1.05, −0.92]. Together, these results suggest that even though phantom rules tested here are also laws, they are psychologically different from more prototypical laws in the U.S. legal system.

4.7. Individual differences about rule preferences

Following our pre-registration, we also explored TLC and RWA as moderators of the effects on the collapsed moral relevance variable and the legitimacy rating. However, because TLC had lower reliability than what we pre-registered, we have relegated those findings to the supplement and only report RWA findings here. Here, each model again includes participants as a random effect.[11]

4.7.1. Right-wing authoritarianism

The pattern of results for the two key dependent variables, moral relevance and legitimacy, was nearly identical. We found a significant interaction between RWA and Rule Type for the moral relevance index, $b = 0.27$, $SE = 0.01$, $t(1987) = 19.44$, $p < .001$, $r = 0.40$, *95% CI* [0.28, 0.39]. Simple slopes analysis suggested that individuals low (−1 SD = −1.88) and average in RWA saw the phantom rule violations as considerably less morally relevant than prototypical legal violations ($b = -1.62$, $SE = 0.04$, $t = -44.49$, $p < .001$; $b = -1.12$, $SE = 0.03$, $t = -43.48$, $p < .001$, respectively). While still statistically significant, the

[10] We attempted to fit all models in Experiments 3–5 with random slopes for participant and stimuli, including random intercepts. However, in most cases, the models failed to converge, and in some cases, the ICC for stimuli did not reach the preregistered cutoff of 0.10. As such, we removed the random slopes, leaving by-participant and by-item random intercepts for all models (Singmann and Kellen, 2019) unless the by-item random effect did not reach the cutoff (as is the case for target models in Experiment 3). In those cases, we included only by-participant and note the change in the text. When models with complex random effects did converge (including when the by-stimuli random intercepts were included), the parameter estimates for fixed effects were comparable and patterns of results were unchanged. When stimulus type is entered as a fixed effect, the patterns of results remain unchanged. The overall pattern of results reported for Experiments 3–5 are robust to these different specifications.

[11] Again, if we include stimulus as random effect or include it as a fixed effect, conclusions remain unchanged.

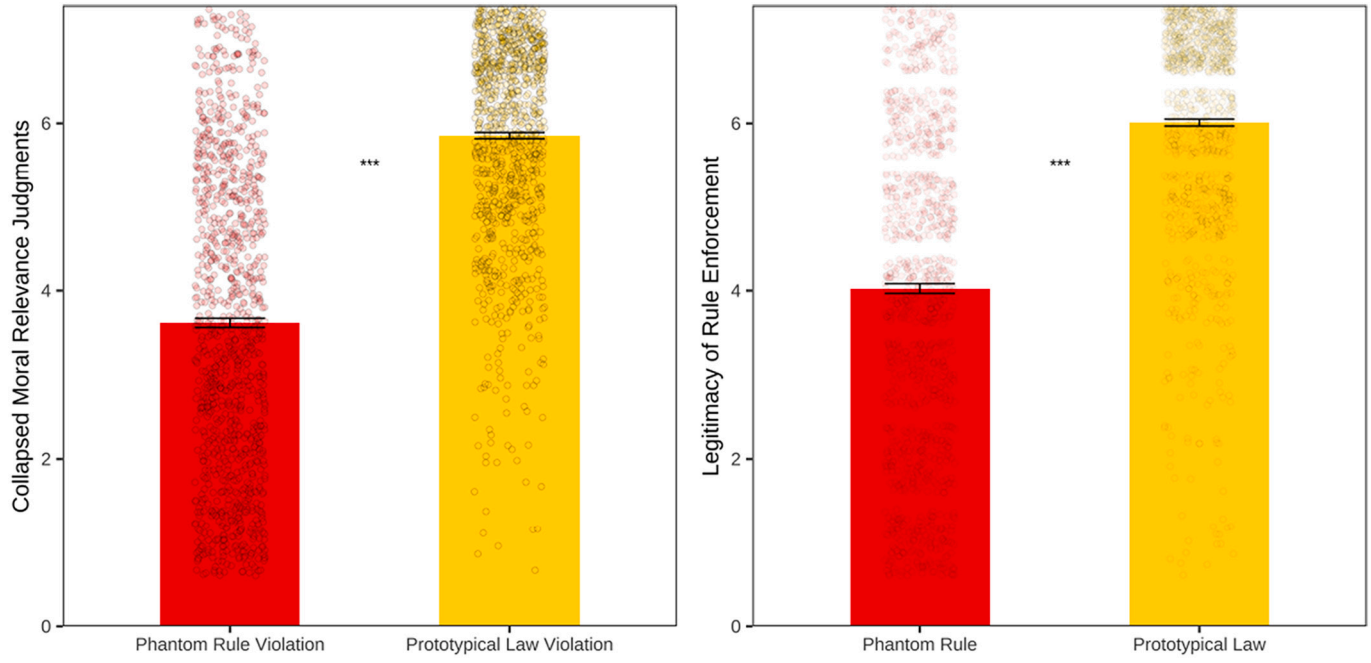

Fig. 5. The collapsed moral relevance variable by violation type (phantom rule and prototypical legal violations) is shown on the left, and legitimacy by rule type (phantom rule and prototypical law) is displayed on the right. '***' corresponds to $p < .001$. Error bars represent ±1 *SEM*. See the online article for the color version of this figure.

magnitude of the slope for those high in RWA (+1 *SD*) was much lower, suggesting that individuals high in RWA see legal rules as more psychologically similar to phantom rules ($b = -0.62$, $SE = 0.04$, $t = -16.99$, $p < .001$), though they still distinguish between them. This pattern held true for legitimacy of enforcement as well, $b = 0.25$, SE = 0.02, t(1987) = 15.71, p < .001, $r = 0.33$, *95% CI* [0.22, 0.29]. The slopes for individuals low and average in RWA were greater in magnitude ($b = -1.47$, $SE = 0.04$, $t = 34.19$, $p < .001$; $b = -0.99$, $SE = 0.03$, $t = 32.64$, $p < .001$, respectively) compared to those high in RWA ($b = -0.51$, $SE = 0.04$, $t = -11.97$, p < .001; see Fig. 6). These results suggest that, as predicted, the legitimacy of phantom rule enforcement and moral relevance of phantom rule violations is significantly lower than that of prototypical laws and their violations—especially for those low in Right Wing Authoritarianism.[12]

4.8. Discussion

We found that, when compared to violations of more prototypical laws, phantom rule violations are judged to be less illegal, less morally relevant, and the enforcement of the rules themselves less legitimate. And, as predicted, they are also judged to be more descriptively normative to break. In an experiment reported in the supplemental material (Experiment S1), we also found that phantom rule violations psychologically differ from non-codified, but socially consequential social norms (e.g., cutting in line). Together, these findings suggest that phantom rule violations are psychologically different from other kinds of violations (both legal and social), including those that are more prototypical laws governing similar behaviors. Phantom rules are codified rules, but they are both descriptively normative and less morally consequential to break. Moreover, we observed the overall predicted pattern, people who are low in RWA see the difference between phantom rules and other rules and laws as meaningful. Although people high in RWA also recognize a small difference between these rules and their violations, this effect is much more pronounced for individuals who are less concerned with following rules and respecting authority.

Phantom rules occupy a special space in our psychology—one that is distinct from both social norms and more prototypical legal rules. Unlike social norm and legal rule violations, it is descriptively normative and morally inconsequential to break phantom rules. Notably, unlike judgments of phantom rule violations, all of the judgments of prototypical legal violations were near ceiling. Taken together, we find evidence for unique ambiguity in the apparent punishability of phantom rule violations.

More specifically, we suggest that judgments of phantom rule punishability is subject to people's motivation to blame, condemn, and punish for wrongdoing (Alicke, 2000; Ames and Fiske, 2013; Haidt, 2001; Robinson and Darley, 1995). Critically, when it comes to phantom rules, we suggest that extant motivation to punish leads to motivated reasoning about the legitimacy (i.e., punishability) of the rule (Kunda, 1990). Motivated reasoning—the tendency to conform judgments and evaluations to a particular goal based on prior beliefs and desires (Kunda, 1990)—infects cognition of all types, including the moral domain (Bandura, 1999; Kahan, 2013; Paharia et al., 2013; Uhlmann et al., 2009). We argue that phantom rules are invoked in a motivated way when people want to punish the transgressor for *something*.

Under a motivated reasoning account of phantom rule punishability, people change their view of the rule itself, enhancing its apparent punishability when they want to invoke (the otherwise dormant) rule. When people have an active motive to blame others, phantom rules—and the punishment that follows from violating them—may then be seen as more justifiable than when this motive is not activated. Phantom rules become a means to enact the desired punishment. If this is the case, we expect to see that, in the absence of a clear avenue for punishment after a perceived violation (e.g., when a social norm is violated), people will judge phantom rules to be more punishable, more legitimate, and their violation to be worse, than when there is no other reason to blame or punish the actor. Indeed, this framework is consistent with other theories of blame that suggest the act of ascribing blame carries with it consequences for subsequent moral judgments, including intentionality,

[12] All main analyses reported above remain when entering RWA as a covariate in the model

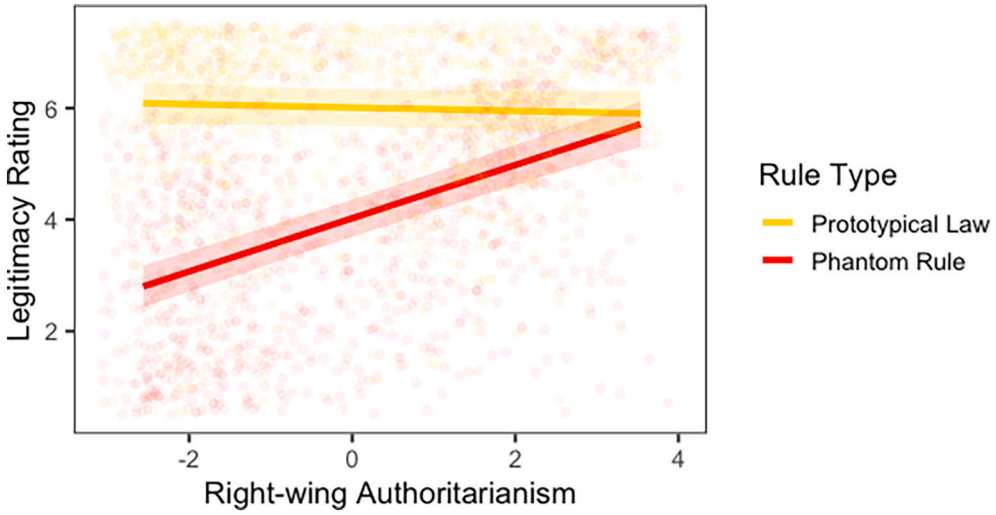

Fig. 6. Interaction between rule type (phantom rule vs. prototypical law) and RWA (centered) for legitimacy ratings. Individuals high in RWA rated phantom rules as similarly legitimate to prototypical laws. This pattern of results was consistent for the moral relevance index as well. Shaded regions represent 95% Confidence Intervals. See the online article for the color version of this figure.

causality, and more (e.g., Alicke, 2000; Knobe, 2003, but see Malle et al., 2014). For example, research has found that a gun possession charge is judged to be morally worse and more intentional when the transgressor was caught while trying to hide cocaine compared to an anniversary present (Nadler and McDonnell, 2011). Here, the desire to punish affects the moral judgments we assume precede and inform "subsequent" punishment judgments. On this view, phantom rules become justifiable to punish only after a motivation to punish brings them from obscurity—not the other way around. Indeed, the motivation to punish may be a pre-requisite for a phantom rule to be viewed as legitimately enforceable at all.

5. Experiment 4a

In Experiments 4a and 4b, we sought to directly test a motivated reasoning account of phantom rule punishability. In Experiment 4a, we examined whether an activated motivation to punish would lead to an *increase* in the legitimacy and justifiability of punishment for the phantom rule itself. Specifically, we predicted that phantom rules are seen as more legitimate to enforce when they are violated by a person who has also violated a social norm than when they are violated in isolation. The social norm violations are not illegal, and so the means to punish someone for violating them are less clear and more personally costly. And, given the diagnostic value of social norm violations (e.g., Ditto and Jemmott, 1989; Fiske, 1980), and dominance of moral judgments in impression formation (Brambilla et al., 2011; Goodwin et al., 2014; Landy et al., 2016; Leach et al., 2007; but see Abele and Wojciszke, 2007 for alternative view), we expected social norm violations to activate punishment motivations. Critically, to test that it is indeed a function of motivated reasoning, we measured whether people judge the phantom rule *itself* to be more deserving of punishment, in addition to the person who violated the rule. The rule remains the same in both conditions, but the motivation to invoke the rule varies. The motivated reasoning follows this directional logic: "I want to punish this person for their social norm violation, they *did* technically also break this phantom rule. That rule should be followed!" Experiment 4a examined the interplay between phantom rule violations alone and those committed in conjunction with a social norm violation.

Table 4
Examples of the vignettes and faces used in Experiment 4a.

Phantom Rule Violation Alone	Phantom Rule + Social Norm Violation
Imagine you are sitting in a park on a day when the weather is nice. You see **Dylan** (pictured above) try to politely ask a stranger for the time. After that, you see that Dylan is sitting on a bench in the park using marijuana.	Imagine you are sitting in a park on a day when the weather is nice. You see **Dylan** (pictured above) trying to provoke a stranger. After that, you see that Dylan is sitting on a bench in the park using marijuana.

Full list available on OSF and in the Supplemental Material.

5.1. Method

5.1.1. Participants

Data were collected from 207 participants using Prolific (118 females, 82 males, 4 other, 75.4% White, 17.9% People of Color, 6.7% Other or did not report, M_{age} = 33.07, SD_{age} = 11.28). Consenting participants were compensated $1.65 for their time. An a priori power analysis was conducted using PANGEA (Westfall, 2016) and it revealed that data from ~180 participants will have 80% power to detect an effect size of d = 0.20, which was decided by rounding up the smallest target effect of interest from in Experiment S1 (reported in Supplemental Material). Given that analyses use linear mixed-effects models, the estimate approximates Cohen's d and was calculated using the suggested method from Westfall and colleagues (Westfall et al., 2014). Since removing participants who did not complete the survey did not change patterns of results, we opted to keep all participants in the sample (as pre-registered).

5.1.2. Design

A within-subjects factor (Rule Type) with two levels (phantom rule violation alone vs. social norm violation and phantom rule violation) was used to investigate whether previous social norm violation heightened judgments of legitimacy of punishment and blameworthiness for phantom rule violations. Participants read a total of six vignettes (6 total blocks) that were paired with a face and common name, one at a time. No faces, names, or scenarios repeated, and the block order and type were randomly presented to participants.

5.2. Materials & procedure

5.2.1. Manipulation/Stimuli

Participants viewed six total vignettes that were paired with a face (to make the task more engaging) in random order, followed by a series of questions measuring their judgments of each vignette. The faces were

all White males selected from the Chicago Face Database (CFD: Ma et al., 2015) and were matched on the dimension of attractiveness (see Supplemental Material). Each vignette either described the named person committing a social norm violation in addition to the phantom rule violation, or the phantom rule violation alone (see Table 4). The vignettes for the social norm violation were taken from previous research and were matched on moral relevance and negativity[13] (Mende-Siedlecki and Havlicek, n.d.). The same person was always paired with a phantom rule violation (regardless of whether a social norm violation was also present) and participants never saw the same phantom rule violation more than once.

5.2.2. Moral & punishment judgments

The primary goal of this experiment was to assess differences between the punishability of phantom rule violations when they are committed in isolation or in conjunction with a more morally consequential (i.e., social norm) violation.

5.2.2.1. Punishment judgments. Two punishment items were assessed, which consisted of: (1) financial punishment ("How justified is it to be fined for [rule violation]?"); and (2) police punishment ("How justified is it to for the legal system to be involved when someone [rule violation]?") rated from 1 = *Not at all justified* to7 = *Completely justified*. We also included one dichotomous judgment ("Imagine a police officer notices [name and violation], do you think the police officer should initiate an interaction?").

5.2.2.2. Moral judgments. The two target moral judgments included one moral action judgment ("How morally bad is [phantom rule violation]?") item rated from 1 = *Not at all morally bad* to 7 = *Extremely morally bad*; and one moral character judgment ("How morally bad is [name from vignette]?") item rated from 1 = *Not at all morally bad* to 7 = *Extremely morally bad*. Participants also made a blame judgment, ("How deserving of blame is [name]?") rated from 1 = *Not at all* to 7 = *Very much*.

5.2.3. Demographics

Participants were asked to indicate their gender, age, race, religion, and political orientation as in Experiment 3.

5.3. Procedure

The procedure was the same as that used in Experiment 3. Consenting participants first saw a randomly presented block with a vignette and an image. They then answered questions about punishment, blame, and morality.[14] After rating each of the six blocks of scenarios, participants then completed questions about demographic information and then were debriefed and compensated. All materials are available on the Open Science Framework.

5.4. Results

We used the same analysis plan as in Experiments 3, using linear mixed-effects models that specify random intercepts for participants and stimuli. In contrast to Experiment 3, only a few models did not meet the pre-registered ICC cut-off of 0.10 and for those models, we note the deviation and drop the stimulus random effect. We again investigated whether the conceptually similar dependent variables could be collapsed into an index (as pre-registered). The two punishment judgments and the blame judgment were collapsed (Cronbach's $\alpha = 0.85$). The other dependent variables were handled as separate for all subsequent analyses for conceptual purposes. All analyses were conducted using the R statistical analysis software (R Core Team, 2019). Correlations between the dependent variables are included in the supplementary material (see Table S6 in the Supplementary Material).

5.4.1. Main analyses

5.4.1.1. Punishment & blame[15]. Using the collapsed variable for punishment and blame we tested for differences between phantom rule violations alone compared to social norm and phantom rule violations together. Critically and as predicted, phantom rule violations paired with social norm violations were rated as more justifiable to punish compared to phantom rule violations alone, $b = 0.29$, $SE = 0.05$, $t(1128) = 6.27$, $p < .001$, $r = 0.18$, *95% CI* [0.20, 0.38] (see Fig. 7).

We then investigated whether the condition influenced choices in the binary judgment of whether the police should initiate an interaction (i.e., punish) the person or not initiate an interaction. To assess a binary outcome in a mixed effect model, we utilized the "glmer" function from the "lmer" package. Results suggested that people more often selected the "Punish" option (as compared to the "No punish") for phantom rule and social norm violations together compared to phantom rule violations alone, $b = 0.32$, $SE = 0.06$, $z = 5.17$, $p < .001$, $r = 0.09$, *95% CI* [0.20, 0.45], $OR = 1.38$. Notably, when a phantom rule and social norm are violated in tandem, participants think the police should initiate an interaction more often, but when the phantom rule is violated alone, participants are more likely to think the police should not intervene.

5.4.1.2. Moral judgments. We next examined whether the phantom rule violation alone compared to social norm and phantom rule violations together influenced judgments of the morality of the particular phantom rule violation and of the person. Results revealed that when a phantom rule was committed following a social norm violation, the phantom rule violation *itself* was seen as more morally wrong compared to when a phantom rule violation was committed alone, $b = 0.20$, $SE = 0.04$, $t(1115) = 4.58$, $p < .001$, $r = 0.14$, 95% CI [0.11, 0.28]. This was also true for the moral character variable: Violators of phantom rules paired with social norm violations were seen as more morally bad than violators of phantom rules alone, $b = 0.52$, $SE = 0.04$, $t(1129) = 12.40$, $p < .001$, $r = 0.35$, *95% CI* [0.11, 0.28].

5.5. Secondary analyses

Here we have evidence for increased judgments of both the punishability of phantom rules and the moral character of the person who violated them. We hypothesized that the punishability of phantom rules is ambiguous, and subject to motivated reasoning when a punishment motive is activated; people want to enact a punishment and so an available avenue for punishment appears more legitimate than it would otherwise. Yet, it is also possible that ascriptions of bad moral character led to a "pragmatic" inference in which phantom rules serve to prevent future violations by bad actors. To begin to adjudicate between these two interpretations, we also explored how each predictor faired in predicting the outcome of the binary punishment choice variable. When the moral wrongness of the action and the moral character of the actor ($r = 0.65$) are entered simultaneously as predictors,[16] the odds ratio for

[13] The valence dimension was rated on a scale from +3 to −3, where −3 is "very negative" and the moral relevance dimension was rated on a scale from 1 to 9, where 9 is "very much related to morality". The selected social norms had a negativity mean of −1.8 and a moral relevance mean of 5.7.

[14] We also measured competence and individual probability of punishment for that scenario, which are beyond the scope of the current paper.

[15] These models exclude stimuli as a random effect due to an ICC below the preregistered cut-off. When including the random effect of stimuli or the fixed effect of stimuli, results remain unchanged.

[16] This model did not include a random effect of stimuli because the ICC was below the pre-registered cut-off. Conclusions remain unchanged when including it as a random or fixed effect.

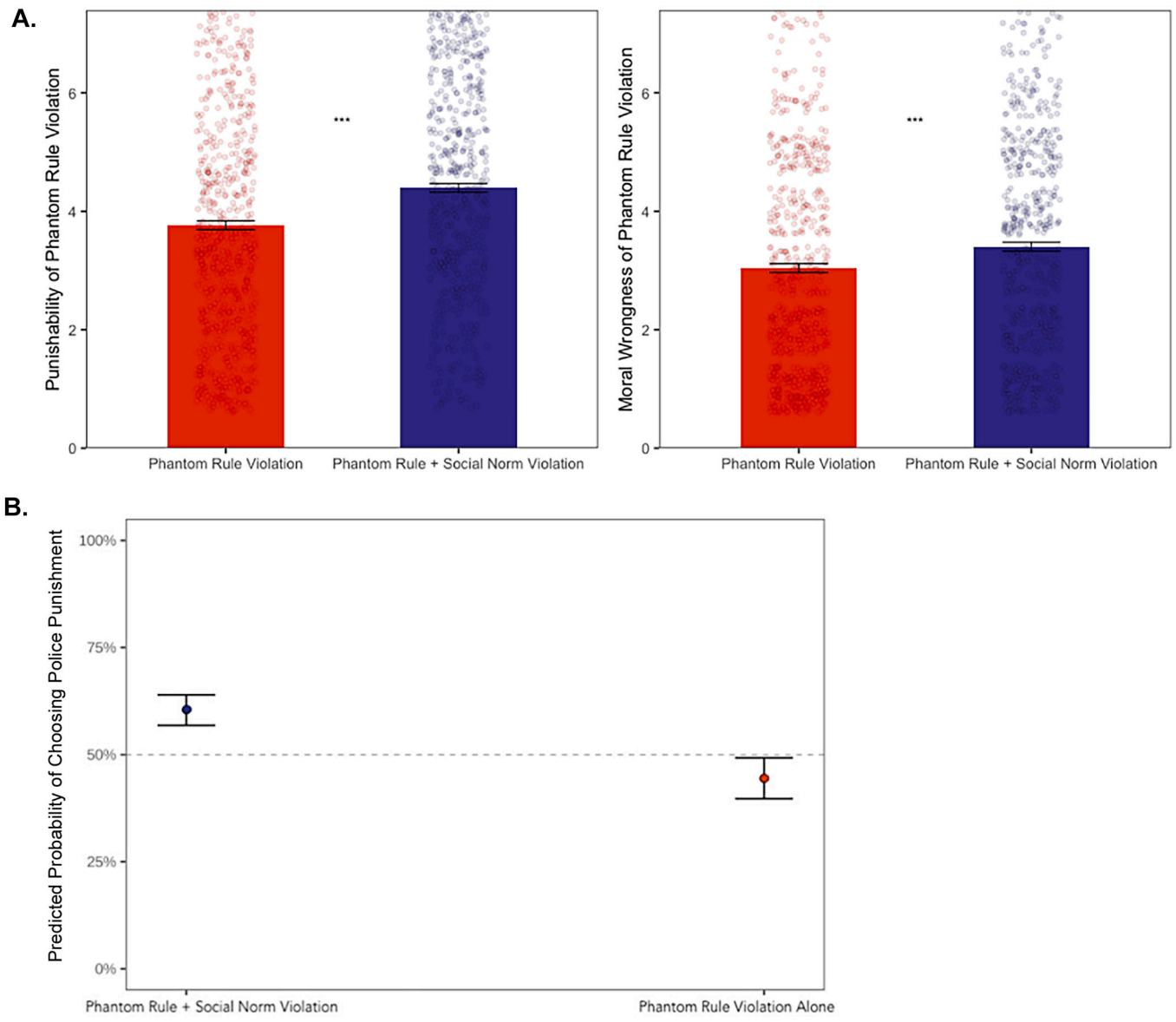

Fig. 7. Moral and punishment differences for phantom rule and social norm violations compared to phantom rule violations alone. (A) Participants find phantom rule violations that were paired with social norm violations more punishable and morally wrong than phantom rule violations alone, even when asked explicitly to rate the rule itself. '***' corresponds to $p < .001$. Error bars represent ±1 *SEM*. (B) The graph shows the predicated probabilities for the fixed effect Scenario Type with a dotted line at 50% (the task is binary). There is an increased probability of selecting to punish the phantom rule (as opposed to letting the violator go unpunished) for the scenarios when a phantom rule violation was paired with a social norm violation (predicted probability = 60%), then when it was alone (predicted probability = 44%). The error bars represent 95% *CI*s and are log-transformed. See the online article for the color version of this figure.

moral wrongness of the action, not the moral character judgment is more strongly associated with the choice to punish, $OR_{action} = 1.80$, *95% CI*$_{action}$ [1.64, 2.09], $OR_{person} = 1.14$, *95% CI*$_{person}$ [1.04, 1.30]. These results begin to suggest that judgments about the wrongness of a phantom rule violation are malleable according to motivation, and lead to a greater probability of selecting to punish the transgressor compared to character judgments.[17]

5.6. Discussion

We found evidence that very same rule is more punishable when a person violated the phantom rule and a social norm together, compared to the phantom rule violation alone. Participants rated the *action itself* as more punishable, and specifically that the police should be involved in a phantom rule violation, when they were motivated to punish the person for a different violation. Notably, in the absence of a social norm violation, people more often than not (below 50%) say they do not want the police to be involved after a phantom rule violation alone. To test whether this effect is special to phantom rules (vs other similar legal violations), we next compared the motivated enforcement of phantom rule violations committed alone compared to those committed in tandem with a social norm violation to that same combination for prototypical legal violations.

6. Experiment 4b

In Experiment 4b, we sought to establish discriminant validity. We

[17] Model comparison suggests that the AIC is lower (better fitting) for the moral action compared to the moral person predictor, suggesting that moral action is a more important predictor.

examined whether an activated motivation to punish would lead to an *increase* in the justifiability of punishment and blame for phantom rule violations but not for prototypical legal violations (like those used in Experiment 3). Experiment 4b used the same logic of Experiment 4a and allowed us to test whether phantom rule violations are particularly susceptible to motivated reasoning in their enforcement. We also included a measure of frustration; we thought that another distinguishing feature of phantom rules from prototypical legal rules is that it is particularly frustrating to be singled out for violating a phantom rule, since most people do so without consequence.

6.1. Method

6.1.1. Participants

Data were collected from 310 participants using Prolific (170 females, 133 males, 7 other, 62% White, 34% People of Color, 4% Other or did not report, $M_{age} = 33.5$, $SD_{age} = 12.2$). Consenting participants were compensated $2.27 for their time. To reach an analyzable sample of at least 280, data was collected from 325 participants (assuming a ~ 15% attention check failure rate), 15 people did not complete any study measures and were not included in analyses. This number was pre-registered and selected to mirror the powering of the previous experiment that follows a similar design with an additional 100 participants to account for differences in the stimuli.

6.1.2. Design

Two within-subjects factors Scenario Type (phantom rule violation alone vs. social norm violation and phantom rule violation) and Rule Type (phantom rule vs. prototypical legal violation) were used to investigate whether previous social norm violation heightened judgments of legitimacy of enforcement and blameworthiness for phantom rule violations and not prototypical legal violations. Following the design of Experiment 4a, participants read a total of six vignettes (6 total blocks) that were paired with a face and common name, one at a time. No faces, names, or scenarios repeated, and the block order and type were randomly presented to participants.

6.2. Materials & procedure

6.2.1. Manipulation/stimuli

We again had participants read six total vignettes that were paired with a face (same as Experiment 4a) in random order. Each vignette either described the named person committing a social norm violation in addition to the phantom rule violation or a prototypical legal violation, or a phantom rule violation or prototypical legal violations alone (see Table 5). All vignettes were taken from Experiment 4a and 3b.

Table 5
Examples of the vignettes and faces used in Experiment 4b.

	Rule Violation Alone	Social Norm + Rule Violation
Phantom rule violation	Imagine you are sitting in a park on a day when the weather is nice. You see **Dylan** (pictured above) try to politely ask a stranger for the time. After that, you see that Dylan is sitting on a bench in the park using marijuana.	Imagine you are sitting in a park on a day when the weather is nice. You see **Dylan** (pictured above) trying to provoke a stranger. After that, you see that Dylan is sitting on a bench in the park using marijuana.
Prototypical legal violation	Imagine you are sitting in a park on a day when the weather is nice. You see **Dylan** (pictured above) try to politely ask a stranger for the time. After that, you see that Dylan is sitting on a bench in the park using cocaine.	Imagine you are sitting in a park on a day when the weather is nice. You see **Dylan** (pictured above) trying to provoke a stranger. After that, you see that Dylan is sitting on a bench in the park using cocaine.

Full list available on OSF and in the Supplemental Material.

6.2.2. Moral & punishment judgments[18]

6.2.2.1. Punishment judgments. We used similar punishment items from Experiment 4a. We asked, "How punishable is the act of [law violation]?" rated from 1 = *Not at all* to 7 = *Very much*, "How justified is it for the legal system to be involved when a person [law violation]?" rated from 1 = *Not at all justified* to 7 = *Completely justified*, and "How blameworthy is the act of [law violation]? Rated from 1 = *Not at all blameworthy* to 7 = *Extremely blameworthy*. We also asked one additional blame item: "How deserving of blame is someone who does [law violation]?", rated from 1 = *Strongly disagree* to 7 = *Strongly agree*. We again included the same dichotomous judgment from Experiment 4a. As in Experiment 4a and in our pre-registration, we collapsed these items into a single index (Cronbach's α = 0.93).

6.2.2.2. Moral judgments. We used the same moral judgment items used in Experiment 4a were used in Experiment 4b with one additional item: "I think [law violation] is indicative of bad character" rated from 1 = *Strongly disagree* to 7 = *Strongly agree*. Given that we had three related items (*r*'s > 0.78), we collapsed across them for analyses (Cronbach's α = 0.92).

6.2.2.3. Legitimacy judgments. We also included five items to assess judgments about the legitimacy and fairness of phantom rules adapted from Tyler et al. (2007) and Tyler and Blader (2005). Representative items include: "How legitimate is the [phantom /prototypical rule violation] law?" rated from 1 = *Not at all* to7 = *Very much*, and "People should follow the laws prohibiting [phantom /prototypical rule violation] even if they think it is wrong." Rated from 1 = *Strongly disagree* to7 = *Strongly agree*. These items had high internal reliability and were collapsed into a single index (Cronbach's α = 0.88).

6.2.2.4. Frustration. We also included three items to assess the affective quality of phantom rule violation enforcement. The items included were, "If you did get punished for [phantom/prototypical rule violation], how frustrated would you be?" rated from 1 = *Not at all Frustrated* to 7 = *Extremely frustrated*, "If you did get punished for [phantom/prototypical rule violation], how unfair would it feel?" rated from 1 = *Not at all unfair* to 7 = *Extremely unfair*, and "If you did get punished for [phantom/prototypical rule violation], how singled-out would you feel?" rated from 1 = *Not at all singled-out* to 7 = *Extremely singled-out*. These items had high internal reliability and were collapsed into a single frustration index (Cronbach's α = 0.86).

6.2.3. Demographics

Participants were asked to indicate their gender, age, race, and political orientation as in Experiment 4a. We pre-registered examining the RWA trait measure, but due to length did not end up including it.

6.3. Procedure

The procedure was the same as that used in Experiment 4a. All materials, including the full Qualtrics survey, are available on the Open Science Framework.

6.4. Results

We used the same analysis plan as in Experiments 4a, all analyses used linear mixed models that specify random intercepts for participants and stimuli (unless otherwise noted). Pairwise differences between marginal means with linear models using the 'emmeans' package (Lenth

[18] We also preregistered and collected data on practical concern judgments. Those analyses are reported in the supplemental material.

et al., 2020). Correlations among the dependent variables are included in the supplementary material (see Table S8 in the Supplementary Material).

6.5. Main analyses

6.5.1. Punishment & blame

We first examined whether Scenario Type (Rule violation alone vs. Rule and a social norm violation) and Rule Type (Phantom rule vs. Prototypical rule), and their interaction term influenced the collapsed punishment and blame judgment index.[19] Results suggested that phantom rule violations lead to less punishment and blame than prototypical legal violations, $b = -2.75$, $SE = 0.09$, t (1773) $= -31.34$, $p < .001$, $r = 0.60$, *95% CI* [−2.92, −2.58]. No significant effect emerged for the Scenario Type, $p = .818$, *95% CI* [−0.20, 0.15]. Contrary to predictions, there was no significant interaction effect, $b = 0.24$, $SE = 0.13$, t (1774) $= 1.82$, $p = .061$, $r = 0.04$, *95% CI* [−0.01, 0.49]. However, when looking to the data in Fig. 8, the patterns support the predicted effects though they should be interpreted with caution. That is, prototypical legal violations are near ceiling, constraining the possible influence of the Scenario Type manipulation.

Following analyses in Experiment 4a, we also tested whether Scenario Type and Rule Type influence the dichotomous punishment judgment. As predicted, phantom rule violations alone were less likely to lead to selection of the punishment option (vs. the "No punish") than prototypical legal violations, $b = -3.26$, $SE = 0.22$, $z = -14.51$, $p < .001$, $r = -0.67$, *95% CI* [−3.71, −2.83], $OR = 0.04$, and rules coupled with a social norm violation were more likely to lead to selection of the punishment than alone, $b = 0.53$, SE $= 0.25$, $z = 2.13$, $p = .033$, $r = 0.14$, *95% CI* [0.05, 1.02], $OR = 1.69$. There was no significant interaction, $p = .75$, *95% CI* [−0.68, 0.48].

6.5.2. Moral judgments

Next, we examined whether Scenario Type and Rule Type affected the collapsed moral judgment index. As predicted, phantom rule violations themselves were rated as less morally wrong than prototypical legal violations, $b = -2.67$, $SE = 0.08$, $t(1735) = -32.83$, $p < .001$, $r = 0.62$, *95% CI* [−2.83, −2.52]. There was no significant difference between the two Scenario Type conditions, $p = .35$, *95% CI* [−0.09, 0.24]. Critically, results revealed a significant interaction, $b = 0.45$, $SE = 0.12$, $t(1736) = 3.83$, $p < .001$, $r = 0.09$, *95% CI* [0.22, 0.69], such that phantom rule violations committed alone were seen as less morally wrong than those committed with a social norm violation, $b = -0.53$, $SE = 0.09$, $t(1750) = -6.26$, $p < .0001$ (see Fig. 8). There was again no statistically significant difference between prototypical legal violations committed alone and those committed alongside a social norm violation, $p = .79$. Violations of phantom rules are specifically subject to motivated rule enforcement.

6.6. Secondary analyses

6.6.1. Legitimacy judgments

Next, we tested whether legitimacy judgments differed between phantom rule and prototypical legal violation judgments based on whether a social norm violation was also committed. Again, phantom rules were rated as less legitimate to enforce than prototypical laws, $b = -1.87$, $SE = 0.08$, $t(1732) = -24.54$, $p < .001$, $r = 0.51$, *95% CI* [−2.02, −1.72], replicating previous patterns. There was no significant effect of Scenario Type and no significant interaction, $ps > 0.14$.

6.6.2. Frustration

Finally, we also asked participants about how frustrated they would feel if they got caught for violating one of these rules. As predicted, participants rated phantom rule violations as more frustrating to get caught for compared to prototypical legal violations, $b = 2.00$, $SE = 0.10$, $t(1677) = 20.91$, $p < .001$, $r = 0.45$, *95% CI* [1.81, 2.19].[20] While, no significant effect emerged for Scenario Type, $p = .588$, *95% CI* [−0.14, 0.24], a significant interaction emerged, $b = -0.33$, $SE = 0.14$, t (1678) $= -2.37$, $p = .018$, $r = 0.06$, *95% CI* [−0.60, −0.06]. Phantom rule violations, especially when committed alone, are rated as more frustrating to get caught for than other rules, $b = 0.28$, $SE = 0.10$, t (1689) $= 16.41$, $p = .028$.

6.7. Discussion

Experiment 4a and 4b demonstrate that phantom rules are particularly susceptible to motivated reasoning about whether or not they should be enforced. We see two primary possibilities for understanding these findings. We hypothesized *a priori* that when people want to punish someone for a social norm violation but do not have an easy way to execute the punishment, phantom rules specifically become more justifiable to punish. This, in essence, is a motivated reasoning account of the apparent punishability of phantom rules. This interpretation is also consistent with work on "just deserts". For example, research in this vein suggests that individuals tend to state preferences for deterrence-based punishment motivations (e.g., reduce probability of repeat infraction), but overwhelmingly demonstrate behavioral preferences for factors associated with just deserts (Carlsmith et al., 2002; see Darley, 2009 for review). Relatedly, work on moral outrage has shown a similarly motivated account of punishment and blame, demonstrating that anger carries over and influences later, unrelated judgments when no compensatory action was taken following an initial crime (Goldberg et al., 1999). But prototypical legal violations are laws that already match our moral intuitions. Punishability of prototypical legal violations alone were already near ceiling, giving little room for the influence of motivation. Thus, these rules were less susceptible to the patterns of motivated enforcement seen with phantom rule violations.

On the other hand, the differences in phantom rule enforcement could be driven by inferences that reflect practical, useful information, akin to a predictability judgment. That is, seeing someone commit more transgressions is informative of their character and potentially diagnostic of future antisocial behavior. From this perspective, the evidenced differences in moral condemnation between conditions could be attributed to reasonable concern about future actions. One way to adjudicate between these opposing interpretations is to compare what happens to moral judgments of phantom rule violations when transgression number is matched, and motivation to punish is satiated.

7. Experiment 5

In Experiment 5, we sought to provide more evidence for the motivated reasoning account of phantom rule punishability. Active motives heighten the salience of relevant means to fulfill that motive (e.g., Kruglanski et al., 2002; Robinson and Berridge, 1993; Zhang et al., 2009). When a motivation to punish is activated, (e.g., by a social norm violation), a phantom rule can be used as a means to enact a system-sanctioned punishment and thereby satiate the motive. Accordingly, we hypothesized that the motivated punishability and legitimacy of phantom rules enforcement would only emerge when the motivation to punish is activated (vs. satiated). As such, we tested whether satiating the initial motivation to punish decreases the perceived punishability

[19] We did not include stimuli as a random effect in this model because it did not meet the preregistered ICC cut-off of 0.10. However, conclusions remained unchanged when including it as a random or fixed effect.

[20] We did not include stimuli as a random effect in this model because it did not meet the preregistered ICC cut-off of 0.10. However, conclusions remained unchanged when including it as a random or fixed effect.

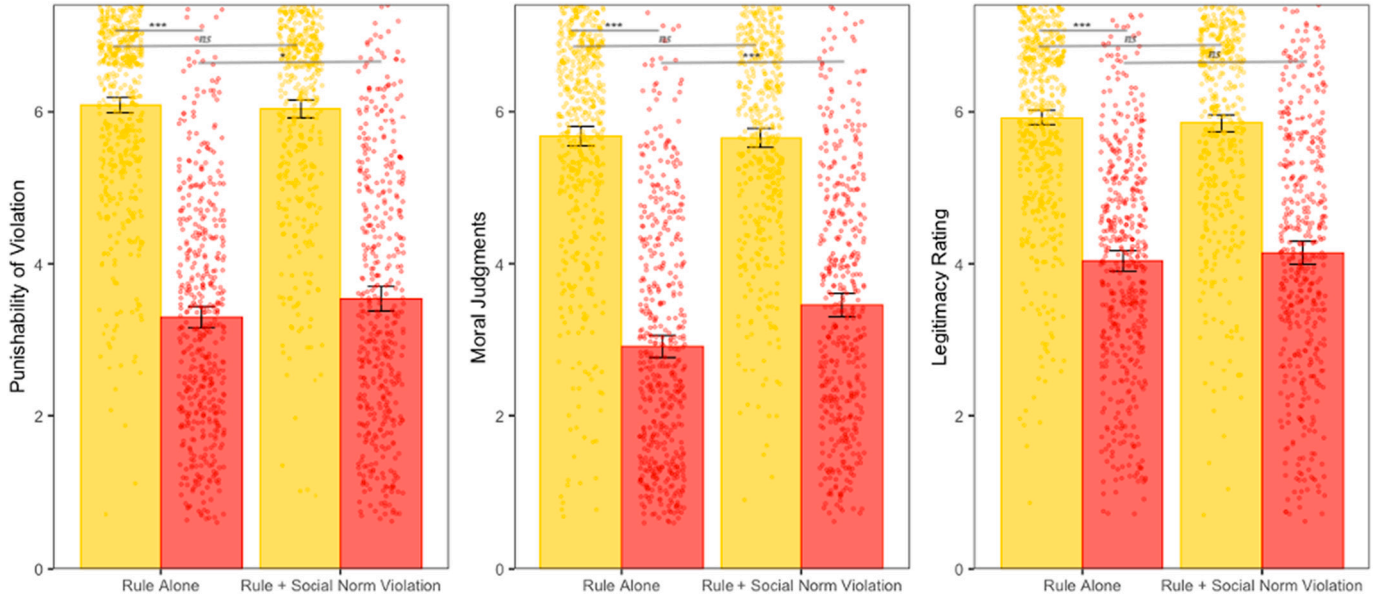

Fig. 8. Differences between phantom rule and prototypical legal violations for punishability, moral wrongness, and legitimacy of enforcement judgments. Participants find prototypical legal violations more punishable, morally wrong, and legitimate to enforce and less susceptible to effects of the social norm manipulation than phantom rule violations. '***' corresponds to $p < .001$, '*' corresponds to $p = .05$, 'ns' corresponds to $p > .06$. Error bars represent ±1 *SEM*. See the online article for the color version of this figure.

and legitimacy of phantom rules compared to when the motive is still active. We tested whether satiating the punishment motive via an intervention targeting the *social norm* violation decreases judgments of the punishability of *phantom rule* violations. Critically, across conditions people read about a person who has committed the same number of violations and all information pertaining to the phantom rule violation itself was identical in both conditions.

7.1. Method

7.1.1. Participants

Data were collected from 285 participants using Prolific (119 females, 153 males, 8 other, 5 did not report, 69.6% White, 28.1% People of Color, 2.3% Other or did not report, $M_{age} = 32.04$, $SD_{age} = 11.34$). Consenting participants were compensated $3.50 for their time. We conducted a simulation a priori to determine if previous recommendations (1600 observations) for mixed-effects models would be sufficient for this design (see Brysbaert and Stevens, 2018). The power analysis utilized parameters from the moral action effect in Experiment 4a since that effect is most similar to the critical effect for the present experiment; it suggested that with 280 participants (1600 observations), power for this design would fall within the interval [83.16, 100.0]. Following our pre-registered data exclusion plan, we removed any participant who failed attention check items, leaving us with a final sample size of $N = 263$ (112 females, 140 males, 5 non-binary, 3 other, 3 did not report, 68% White, 28.9% People of Color, 3.1% Other or did not report, $M_{age} = 31.79$, $SD_{age} = 11.01$).

7.1.2. Design

A within-subjects factor (punishment motivation) with two levels (activated vs satiated) was used to investigate whether satiating the motivation to punish decreases judgments of legitimacy of phantom rules and their enforcement. Participants again read six vignettes paired with a face and common name, one at a time. No faces, names, or scenarios repeated, and the block order and motive type were randomly presented to participants.

7.2. Materials & procedure

7.2.1. Manipulation/Stimuli

The same stimuli from Experiment 3 were used in Experiment 5 with one exception, for the satiated motivation condition, participants read an extra sentence about a third-party reprimanding the social norm violation (i.e., a social punishment for a social violation) that included acknowledgement of the wrongdoing by the transgressor (see Table 6). We included an expression of remorse to ensure that participants believed that the reprimand was effective, and ensure that the punishment motive evoked by the story was satiated.

7.2.2. Moral & punishment judgments

The primary aim was to assess the differences between active compared to satiated punishment motivation on the apparent punishability of phantom rule violations.

7.2.2.1. Punishment judgments. The moral judgments items matched those used in Experiment 4b, including the item to assess deservingness of punishment ("The act of [phantom rule violation] deserves punishment"), rated from 1 = *Strongly disagree* to7 = *Strongly agree*. As preregistered, and because of the high internal reliability of these items, we collapsed them into a single index, (Cronbach's $\alpha = 0.90$).

Table 6
Examples of the vignettes and faces used in Experiment 5.

Active Motivation	Satiated Motivation
Imagine you are sitting in a park on a day when the weather is nice. You see **Dylan** (pictured above) trying to provoke a stranger. After that, you see that Dylan is sitting on a bench in the park using marijuana.	Imagine you are sitting in a park on a day when the weather is nice. You see **Dylan** (pictured above) trying to provoke a stranger. After that, you see that Dylan is sitting on a bench in the park using marijuana. Moments later, you see a person nearby sit down on the bench and criticize Dylan for his provocation attempt. He looks remorseful.

Full list available on OSF and in the Supplemental Material.

7.2.2.2. Moral judgments. Both the moral action and moral character items matched those of Experiment 3. We also included one additional face-valid item to assess moral character ("I think [phantom rule violation] is indicative of bad character"), which was rated on a scale from 1 = *Strongly disagree* to 7 = *Strongly agree*. These items were investigated individually because of low internal reliability and conceptual distinctiveness.

7.2.3. Other judgments[21]

7.2.3.1. Legitimacy judgments. We also included items to assess judgments about the legitimacy and fairness of phantom rules. For legitimacy, we used one item to assess the importance of having a phantom rule: "Should the legal system consider getting rid of the law prohibiting [phantom rule]?" rated from 1 = *Strongly disagree* to 7 = *Strongly agree* (reverse-coded). The other items were the same items that were used in Experiment 4b. Based on high internal reliability, these items were collapsed into a single index (Cronbach's $\alpha = 0.87$).

7.2.4. Demographics

We included the same demographic questions from Experiment 4b.

7.2.5. Procedure

The procedure was the same as that used in Experiment 4a and 4b.

7.3. Results

We used the same analysis plan as in Experiments 3–4, utilizing linear mixed-effects models for analyses with random intercepts for participants and stimuli (unless otherwise noted). Conceptually similar dependent variables were collapsed (Punishment Cronbach's $\alpha = 0.90$) and all other dependent variables were handled individually. Correlations between the dependent variables and exploratory moderators are included in the supplementary material (see Table S10 and S11 in the Supplementary Material for correlations).

7.4. Main analyses

7.4.1. Punishment & blame judgments

Using the collapsed variable for punishment and blame we tested for differences between activated and satiated motivation to punish (i.e., unpunished vs punished social norm violation).[22] Critically, and as predicted, the active motivation condition yielded higher ratings of punishability compared to the satiated motivation condition, $b = 0.08$, $SE = 0.04$, $t(1419) = 2.10$, $p = .036$, $r = 0.06$, *95% CI* [0.01, 0.15] (see Fig. 9).

Additionally, we again tested whether the condition influenced choices in the binary judgment of whether the police should initiate an interaction with the person, following the same analysis strategy used in Experiment 3. As hypothesized, results suggested that people more often selected the "Punish" option for the active motivation condition compared to the satiated motivation condition, $b = 0.22$, $SE = 0.06$, $z = 3.41$, $p < .001$, $r = 0.06$, *95% CI* [0.09, 0.35], $OR = 1.24$. Active but not satiated punishment motivation leads participants to say that police involvement is warranted.

7.4.2. Moral judgments

Next, we tested whether activated punishment motivation compared to satiated punishment motivation would influence judgments of moral wrongness of the phantom rule violation and moral character separately. Results supported predictions; when punishment motivation was active (i.e., the social norm violation went unpunished), the phantom rule violation itself was rated more morally wrong compared to the satiation condition, $b = 0.10$, $SE = 0.04$, $t(1403) = 2.68$, $p = .007$, $r = 0.07$, *95% CI* [0.03, 0.18]. This was also true for the moral character variables: When there was an active punishment motivation (i.e., the social norm violation went unpunished), violators of phantom rules were seen as more morally bad and had worse moral character than violators judged in the satiated condition, $b_{moralbad} = 0.17$, $SE = 0.04$, $t(1422) = 4.63$, $p < .001$, $r = 0.12$, *95% CI* [0.10, 0.25][23]; $b_{character} = 0.11$, $SE = 0.04$, $t(1411) = 2.89$, $p = .004$, $r = 0.08$, *95% CI* [0.03, 0.18] (see Fig. 9).

7.4.3. Legitimacy judgments

We also explored the extent to which an active punishment motivation influenced beliefs about the legitimacy of phantom rules enforcement. Critically, for the active motivation to punish, the phantom rules were rated more legitimate to enforce compared to the satiated condition, $b = 0.08$, $SE = 0.03$, $t(1421) = 2.36$, $p = .018$, $r = 0.06$, *95% CI* [0.01, 0.14], suggesting that having an active motivation to punish *increases* the perceived legitimacy of phantom rules. This is consistent with the idea that phantom rules are invoked as a means to satisfy punishment motivations.

7.5. Discussion

We found evidence that phantom rules are judged to be more punishable when raters have active, but not satiated punishment motives. Notably, the punishment motivation (and satiation) pertains to a social norm violation (not the phantom rule violation itself), and we find this effect when the total number of transgressions was matched across conditions. In keeping with Goal Systems Theory (Kruglanski et al., 2002, see also Fishbach and Ferguson, 2007) we find that the phantom rule is judged more legitimate as a means to satiate the punishment motive invoked by the social norm violation—only when the motive remains activated and is not satiated in some other way. When the motivation to punish was satiated by punishing the *social norm violation*, phantom rule violations were less likely to be rated as morally consequential or warranting police intervention. People use the codified rules as a way to punish behavior that is outside the purview of those rules—with potentially serious consequences, including police involvement.

We opted to use a reprimand to satiate the punishment motive in this study. To ensure that participants thought that the reprimand was effective, we included that the target expressed remorse after being reprimanded. We recognize two issues with this choice. One is that the reprimand came after the phantom rule violation, and so could it be interpreted as a reprimand for both the social norm and the phantom rule violation. Another is that looking remorseful has implications for judgments of the person's character. Indeed, it is possible that the transgressor expressing remorse is the reason that we see differences in character judgments in this study. Importantly, we would not expect expressing remorse for breaking a social norm to affect the punishability of the phantom rule, and we find that rule judgments predict punishment judgments better than character judgments, though both have predictive power. Finally, changes in ratings of the legitimacy of phantom rules by motivation lend additional support to the idea that motivations change judgments of the rules themselves, in addition to judgments of the person who broke them.

[21] We measured judgments about practical concern and report those in the supplementary materials.

[22] We did not include stimuli as a random effect in this model because it did not meet the preregistered ICC cut-off of 0.10. However, conclusions remained unchanged when including it as a random or fixed effect.

[23] This model does not include stimuli as a random effect because it did not meet the pre-registered ICC cut-off.

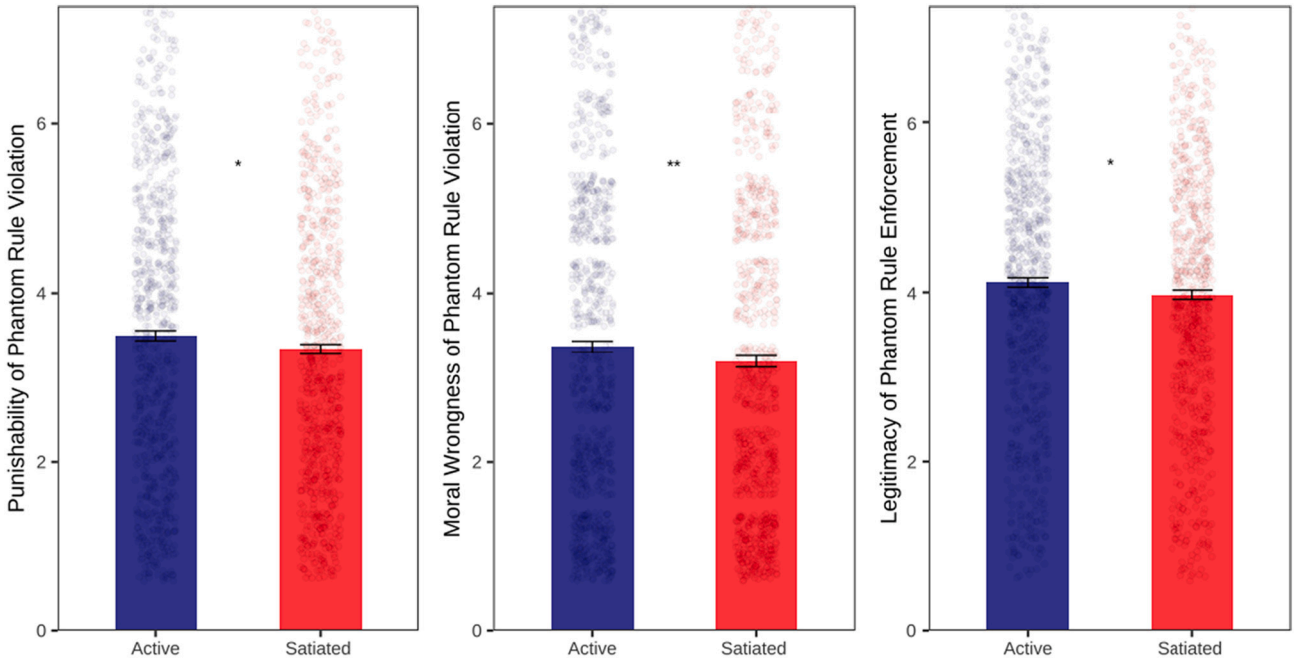

Fig. 9. Punishment, moral wrongness, and legitimacy differences by motivation condition for phantom rule violations. '*' corresponds to $p < .05$, '**' corresponds to $p < .01$. Error bars represent ± 1 *SEM*. See the online article for the color version of this figure.

8. General discussion

In this paper, we identified a subset of rules, which are explicitly codified (e.g., in professional tennis, in an economic game, by the U.S. legal system), frequently violated, and rarely enforced. As a result, their apparent punishability is particularly ambiguous and subject to motivation. These rules show us that codified rules, which are meant to apply equally to all, can be used to sanction behaviors outside of their jurisdiction. We named this subclass of rules phantom rules and found evidence that people enforce them according to their desire to punish a different behavior (i.e., a social norm violation), recognize them in the U.S. legal system, and employ motivated reasoning to determine their punishability. We hypothesized and found, across behavioral and survey experiments, that phantom rules—rules where the descriptive norms of enforcement are low—seem enforceable, punishable, and legitimate only when one has an external active motivation to punish. Indeed, we found that phantom rules were judged to be more justifiably enforced and more morally wrong to violate when the person who broke the rule had also violated a social norm—unless they were also punished for that social norm violation. Together, we take this as evidence of the existence of phantom rules and the malleability of their apparent punishability via active (vs. satiated) punishment motivation.

The ambiguity of phantom rule enforcement makes it possible for them to serve a hidden function; they can be used to punish behavior outside of the purview of the official rules. Phantom rule violations are technically wrong, but on average, seen as less *morally* wrong. This means, for the most part, that people are unlikely to feel strongly when they see these rules violated, and indeed, people frequently violate phantom rules without consequence. This pattern fits well with previous work in experimental philosophy that shows that motivations can affect how we reason about what constitutes breaking a rule in the first place. For example, when blameless rule breaking occurs (e.g., unintentionally), people are less likely to say a rule was violated at all and look for reasons to excuse their behavior (Turri, 2019; Turri and Blouw, 2015). Indeed, our findings mirror this pattern. People find a reason to punish phantom rule violations only when people are particularly or dispositionally motivated to punish. Further, research on indeterminate rules, or rules that are not sufficiently articulated, also points to a key role for motivation in how people reason about whether a rule has been broken (or is punishable). This work finds that when people feel empathy for the rule transgressor, judgments of rule violations decrease (LaCosse and Quintanilla, 2021). That is, like phantom rules, indeterminate rules allow for motivations to shape conceptions of the rules in terms of whether they were broken (LaCosse and Quintanilla, 2021) and whether should be enforced (as in the present research). However, rather than being merely underdetermined, phantom rules contain a unique tension between two sources of information—codified rule information and descriptive norm information. Critically, since phantom rules are technically legitimate to invoke, they can provide a relatively easy avenue for punishment (unlike costly direct punishment; see Balafoutas et al., 2016). Critically, the patterns of phantom rule enforcement presented here push up against the common intuition that legal rules do (and should) apply equally to all (Hannikainen et al., 2021).

Another way to understand our findings is through the lens of moral character judgments. For example, research on person-centered approaches to moral judgments suggests that blame ascriptions are preceded by or totally subsumed by moral character evaluations (Inbar et al., 2012; Knobe, 2010; and see Pizarro and Tannenbaum, 2012 for review). Other research suggests that even deservingness of blame may fundamentally be shaped by character and personal history (Alicke, 2008; Gill and Ungson, 2018). Accordingly, it is possible to interpret the results of Experiments 4a, 4b, and 5 as evidence that people punish others out of practical concern that the person will commit further violations, possibly due to their character. And indeed, in Experiments 4a, 4b, and 5, people rated those who violate both a phantom rule and a social norm together as morally worse people, who are more likely to commit other violations in the future, than those who violated the phantom rule alone. However, even when given the opportunity to tell us about their judgments of the transgressor's moral character, participants still rate the *action* itself as morally worse. And, though both play a role, action (vs. person) ratings better predict whether people think the police should enforce the phantom rule. Thus, while pragmatic concerns about future rule breaking likely play some role in reasoning about punishment of phantom rules, we suggest that people employ motivated reasoning to satisfy their own punishment motives when they think about whether a phantom rule violation warrants punishment or not.

And, regardless of whether the character or the action more strongly contributes to punishment, the law proscribes acts. Laws are not meant to proscribe people or acts outside of their purview.

8.1. Phantom rules and reasoning about the law

The current work makes an important contribution to the understanding of how people reason about codified rules. For example, phantom rules show us that legal rules and moral rules are necessarily dissociable (Hart, 1961), and that this dissociation bears consequences for how much guiding force laws have (e.g., Schauer, 2015). At least in the case of phantom rules, the law itself is not sufficient (unless one is dispositionally inclined to see rules as important) to engender adherence. These findings also reveal important connections with work on how people reason about the letter of the law (i.e., literal meaning) and the spirit of the law (i.e., intention). For instance, research suggests that the spirit of the law accounts for more culpability than its literal meaning (Garcia et al., 2014), and that people sometimes think about the letter and sometimes think about the purpose or intention of a rule (Struchiner, Hannikainen, & Almeida, 2020). It could be, that when people are motivated to punish the rule-breaker, they weigh the letter (vs. the purpose) of the law more heavily as it provides a justifiable, though less common, outlet for punishment (i.e., the person can get them on a "technicality"). This is an interesting avenue for future research, which could investigate whether motivations to punish shift reasoning about the broken rule to the letter of the law from the spirit of the law.

8.2. Limitations

Overall, we find that the perceived punishability and legitimacy of phantom rules is influenced by factors outside of the behavior they proscribe. One limitation of the present analysis is that we did not present evidence that people spontaneously invoke phantom rules, as we suggest the umpire did when he penalized Serena Williams for being coached. It is an interesting question for future research whether motivation to punish enhances the salience of phantom rule violations. Another limitation of the present research is that we only tested phantom rules that are also laws in the United States, where the authors have some working knowledge of what behaviors are illegal. Other cultures and societies may fundamentally differ in how they judge phantom rule violations, or if they even exist at all. Future research should investigate whether other societies differ how they judge phantom rules and whether they are present across cultures.

By necessity, we could only test a subset of phantom rules in real world contexts. Though we validated that our phantom rules met our criteria (they are frequently violated and infrequently enforced), emerged in a controlled economic game, modeled our stimuli as random, and established discriminant validity by comparing them to both legal rules and social norms, we also observed heterogeneity in judgments among them (and so obtained different results when we model the stimuli as fixed, rather than as random effects). We also focused our comparison of phantom rule violations to social norm violations in the legal context, but it would also be interesting to contrast them with violations that occur in other organized systems like professional sports. To address both of these limitations, future research should broaden the scope of phantom rules tested in real world settings and the systems of rules that codify them.

Finally, it is also important to note that no laws or rules have perfect enforcement. There are many reasons why people break laws and then sometimes face consequences, and sometimes break laws and do not. It is no secret that in the United States, that there is massive racial injustice in the criminal justice system for rule violations ranging from jaywalking (Samoylov et al., 2020) to murder (Jacobs et al., 2005; Soss et al., 2003; Zeisel, 1981). Phantom rules are not unique in this sense though we do see evidence for greater susceptibility to motivation compared to more prototypical legal rules in our study that does not take race into account. It is also possible that phantom rules are unique in the sense that people tend not to think that they should be enforced when done in isolation (see Fig. 9 for evidence that people do not want to see police involvement for phantom rule violations alone), unlike more prototypical legal rules which most people tend to say should be enforced (see Fig. 8 for evidence of a ceiling effect for the punishability of more prototypical legal rules). In Experiments 3–4 people are more likely to say the police should not get involved, unless they also saw the person violate a social norm. It may be that phantom rule violations are below the threshold of punishability unless some other reason to punish is present. In other words, the enforcement of phantom rules is infrequent enough, that when they are enforced, they serve a different function than punishment for the violation itself. These rules can function to punish *other* violations. This may also be why being punished for breaking a phantom rule is more frustrating than being punished for breaking a more prototypical law as found in Experiment 3b.

8.3. Future directions

Part of the aim of the present research on phantom rules is to provide a generative framework for this subclass of rules, and so we see a number of interesting avenues for future research. For example, research suggests that statistical norms powerfully influence judgments of blame and praise (Bostyn and Knobe, 2020). Future research should investigate the point at which it is so descriptively normative to break a rule that it becomes a phantom rule—the inflection point at which a codified rule is rendered morally irrelevant.

We focused primarily here on the punishability of phantom rules as motivated reasoning. We think there are many other factors that contribute to the motivated enforcement of phantom rules, and one of the most important factors for future research is the role of identity. For example, people of color received 84% tickets for jaywalking in New York City in 2018 (Samoylov et al., 2020). Indeed, this trend of minority populations being over policed is not a new one, nor is it isolated to New York City. In Alabama, Black individuals are around four times more likely than White individuals to be arrested for marijuana possession even though base rate use is approximately the same (Southern Poverty Law Center, 2018; Thompson, 2017). We expect that people judge the enforcement of phantom rules to be more justifiable when the person who breaks the rule is not White. Research on shifting standards also supports this notion (e.g., Biernat et al., 1991; Biernat and Fuegen, 2001; Biernat and Manis, 1994)—it may be the case that minority populations and women (e.g., Chawla et al., 2020) in the United States see greater enforcement of phantom rules because of shifted expectations of what constitutes normative behavior. In sum, it is a critical future direction of this work to examine the role of race and other social identities in the enforcement of phantom rules, especially in comparison to other more prototypical legal rules.

Another important direction for this work is to investigate the mechanisms that support motivated reasoning about the enforcement of phantom rules. One possibility is that people imagine a more flagrant or egregious violation depending on the person (potentially because of their identity or whatever they were just doing before). For example, it could be that when people read about a man catcalling a woman and then jaywalking, that the instance of jaywalking is egregious (blocking traffic, resulting in lots of honking and screeching, and is generally disruptive and violates the spirit of the law) whereas the instance of jaywalking pictured when the man waves at a woman first is more mild (not causing further disruption and so violates only the letter of the law). It is also possible that people attend to different aspects of the very same scene depending on features of the target or the observer (see Granot, Balcetis, Schneider, & Tyler, 2014).

9. Conclusion

In this paper, we sought to identify and validate the idea of phantom rules, explicit rules or laws, which are frequently violated, rarely enforced, and sometimes invoked to punish norm violations outside of their purview. We found that people enforce these rules more often for selfish (vs. fair) players in an economic game. Then we illustrated that phantom rules are a general subclass of rules, recognizable and distinct other kinds of rules and laws. Critically, when a motivation to punish is activated by some other violation, phantom rule violations are judged to be more legitimate to enforce. These frequently violated rules are invisible to us, like phantoms, unless a motivation to punish the transgressor for some other violation calls them from the shadows.

Author contributions

JW developed the concept, conducted the analyses, contributed to the study design, and contributed to the manuscript. AG contributed to concept development and refinement, study design, and the manuscript.

Data availability

We have included a link to the project's Open Science Framework (OSF) page, which contains materials, code, and data from all reported studies.

Acknowledgements

We thank Matthew Vanaman for helpful conversations that took place in the early stages of this project. We would also like to thank Dave Gantman, Daniel Yudkin, Pia Dietze, and Hannah Nam and the members of the Gantman Lab for their valuable feedback on earlier drafts of this manuscript, and Levi Satter and Z Longo for their help in collecting and cleaning these data. Finally, we would like to thank our anonymous reviewers for their thoughtfulness. This paper is much better for it.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2022.105323.

References

Abele, A. E., & Wojciszke, B. (2007). Agency and communion from the perspective of self versus others. *Journal of Personality and Social Psychology, 93*(5), 751.
Adorno, T. W., Frenkel-Brunswik, E., Levinson, D., & Sanford, R. N. (1950). *The authoritarian personality*. Harper & Row.
Alicke, M. (2008). Blaming badly. *Journal of Cognition and Culture, 8*(1–2), 179–186.
Alicke, M. D. (2000). Culpable control and the psychology of blame. *Psychological Bulletin, 126*(4), 556.
Altemeyer, B. (1981). *Right-wing authoritarianism*. University of Manitoba press.
Altemeyer, B. (1988). *Enemies of freedom: Understanding right-wing authoritarianism*. Jossey-Bass.
Altemeyer, B. (1998). The other "authoritarian personality". In , *Vol. 30. Advances in experimental social psychology* (pp. 47–92). Academic Press.
Ames, D. L., & Fiske, S. T. (2013). Intentional harms are worse, even when they're not. *Psychological Science, 24*(9), 1755–1762.
Arnold, D., Dobbie, W., & Yang, C. S. (2018). Racial bias in bail decisions. *The Quarterly Journal of Economics, 133*, 1885–1932. https://doi.org/10.1093/qje/qjy012
Asch, S. E. (1956). Studies of independence and conformity: I. a minority of one against a unanimous majority. *Psychological Monographs: General and Applied, 70*(9), 1.
Balafoutas, L., & Nikiforakis, N. (2012). Norm enforcement in the city: A natural field experiment. *European Economic Review, 56*, 1773–1785.
Balafoutas, L., Nikiforakis, N., & Rockenbach, B. (2016). Altruistic punishment does not increase with the severity of norm violations in the field. *Nature Communications, 7*, 1–6.
Bandura, A. (1999). Moral disengagement in the perpetration of inhumanities. *Personality and Social Psychology Review, 3*(3), 193–209.
Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., … Bolker, M. B. (2015). Package 'lme4'. *Convergence, 12*(1), 2 (CRAN).
Bearden, W. O., & Etzel, M. J. (1982). Reference group influence on product and brand purchase decisions. *Journal of Consumer Research, 9*(2), 183–194.
Bender, A., Bingham, S., Castaldi, M., Piana, E. D., Besautels, M., Herlad, M., … Zhen, T. (2021, May 4). *Not just a Ferguson problem – How Traffic Courts Drive inequality in California*. LCCRSF. https://lccrsf.org/pressroom_posts/not-just-a-ferguson-problem-how-traffic-courts-drive-inequality-in-california/.
Bernhard, H., Fehr, E., & Fischbacher, U. (2006). Group affiliation and altruistic norm enforcement. *American Economic Review, 96*(2), 217–221.
Bicchieri, C. (2006). *The grammar of society: The nature and dynamics of social norms*. Cambridge University Press.
Bicchieri, C., Muldoon, R., & Sontuoso, A. (2011). *Social norms*. The Stanford Encyclopedia of Philosophy. https://plato.stanford.edu/archives/win2018/entries/social-norms/.
Biernat, M., & Fuegen, K. (2001). Shifting standards and the evaluation of competence: Complexity in gender-based judgment and decision making. *Journal of Social Issues, 57*(4), 707–724.
Biernat, M., & Manis, M. (1994). Shifting standards and stereotype-based judgments. *Journal of Personality and Social Psychology, 66*(1), 5.
Biernat, M., Manis, M., & Nelson, T. E. (1991). Stereotypes and standards of judgment. *Journal of Personality and Social Psychology, 60*(4), 485.
Bostyn, D. H., & Knobe, J. (2020). The shape of blame. *PsyArXiv.*. https://doi.org/10.31234/osf.io/2hca8
Brambilla, M., Rusconi, P., Sacchi, S., & Cherubini, P. (2011). Looking for honesty: The primary role of morality (vs. sociability and competence) in information gathering. *European Journal of Social Psychology, 41*(2), 135–143.
Brauer, M., & Chaurand, N. (2010). Descriptive norms, prescriptive norms, and social control: An intercultural comparison of people's reactions to uncivil behaviors. *European Journal of Social Psychology, 40*(3), 490–499.
Brauer, M., & Chekroun, P. (2005). The relationship between perceived violation of social norms and social control: Situational factors influencing the reaction to deviance. *Journal of Applied Social Psychology, 35*(7), 1519–1539.
Brennan, G., Eriksson, L., Goodin, R. E., & Southwood, N. (2013). *Explaining norms*. Oxford University Press.
Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition, 1*(1).
Buckholtz, J. W., & Marois, R. (2012). The roots of modern justice: Cognitive and neural foundations of social norms and their enforcement. *Nature Neuroscience, 15*(5), 655–661. https://doi.org/10.1038/nn.3087
Buckholtz, J. W., Asplund, C. L., Dux, P. E., Zald, D. H., Gore, J. C., Jones, O. D., et al. (2008). The neural correlates of third-party punishment. *Neuron, 60*(5), 930–940. https://doi.org/10.1016/j.neuron.2008.10.016
Carlsmith, K. M., Darley, J. M., & Robinson, P. H. (2002). Why do we punish? Deterrence and just deserts as motives for punishment. *Journal of Personality and Social Psychology, 83*(2), 284.
Chawla, M., Earp, B. D., & Crockett, M. J. (2020). A neuroeconomic framework for investigating gender disparities in moralistic punishment. *Current Opinion in Behavioral Sciences, 34*, 166–172.
Cialdini, R. B., & Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annual Review of Psychology, 55*, 591–621.
Clarey, C. (2009, November 30). *Probation and fine for Serena Williams's Tirade*. New York Times. Retrieved from https://www.nytimes.com/2009/12/01/sports/tennis/01serena.html.
Clutton-Brock, T. H., & Parker, G. A. (1995). Punishment in animal societies. *Nature, 373*, 209–216.
Crockett, M. J., Özdemir, Y., & Fehr, E. (2014). The value of vengeance and the demand for deterrence. *Journal of Experimental Psychology: General, 143*(6), 2279.
Curtis, M. B., Robertson, J. C., Cockrell, R. C., & Fayard, L. (2021). Peer ostracism as a sanction against wrongdoers and whistleblowers. *Journal of Business Ethics, 174*(2), 333–354.
Darley, J. M. (2009). Morality in the law: The psychological foundations of citizens' desires to punish transgressions. *Annual Review of Law and Social Science, 5*, 1–23.
Ditto, P. H., & Jemmott, J. B. (1989). From rarity to evaluative extremity: Effects of prevalence information on evaluations of positive and negative characteristics. *Journal of Personality and Social Psychology, 57*(1), 16.
Dunlea, J. P., & Heiphetz, L. (2021). Moral psychology as a necessary bridge between social cognition and law. *Social Cognition, 39*(1), 183–199.
Eriksson, K., Strimling, P., & Coultas, J. C. (2015). Bidirectional associations between descriptive and injunctive norms. *Organizational Behavior and Human Decision Processes, 129*, 59–69.
Falk, A., Fehr, E., & Fischbacher, U. (2003). On the nature of fair behavior. *Economic Inquiry, 41*, 20–26.
Falk, A., Fehr, E., & Fischbacher, U. (2008). Testing theories of fairness—Intentions matter. *Games and Economic Behavior, 62*(1), 287–303.
Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149–1160.
Fehr, E., & Fischbacher, U. (2002). Why social preferences matter–the impact of non-selfish motives on competition, cooperation and incentives. *The Economic Journal, 112*(478), C1–C33.
Fehr, E., & Fischbacher, U. (2003). The nature of human altruism. *Nature, 425*(6960), 785–791.
Fehr, E., & Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and Human Behavior, 25*(2), 63–87.
Fehr, E., & Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review, 151*, 867–868. https://doi.org/10.1126/science.151.3712.867-a
Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature, 415*, 137–140. https://doi.org/10.1038/415137a

Fehr, E., Fischbacher, U., & Gächter, S. (2002). Strong reciprocity, human cooperation, and the enforcement of social norms. *Human Nature, 13*(1), 1–25.

FeldmanHall, O., Sokol-Hessner, P., Van Bavel, J. J., & Phelps, E. A. (2014). Fairness violations elicit greater punishment on behalf of another than for oneself. *Nature Communications, 5*, 5306.

Fishbach, A., & Ferguson, M. J. (2007). *The goal construct in social psychology*.

Fiske, S. T. (1980). Attention and weight in person perception: The impact of negative and extreme behavior. *Journal of Personality and Social Psychology, 38*(6), 889.

Flanagan, B., & Hannikainen, I. R. (2022). The folk concept of law: Law is intrinsically moral. *Australasian Journal of Philosophy, 100(1)*, 165–179.

Forsythe, R., Horowitz, J. L., Savin, N. E., & Sefton, M. (1994). Fairness in simple bargaining experiments. *Games and Economic Behavior, 6*(3), 347–369.

Garcia, S. M., Chen, P., & Gordon, M. T. (2014). The letter versus the spirit of the law: A lay perspective on culpability. *Judgment and Decision making, 9*(5).

Gelfand, M. J., Nishii, L. H., & Raver, J. L. (2006). On the nature and importance of cultural tightness-looseness. *Journal of Applied Psychology, 91*(6), 1225.

Gelfand, M. J., Raver, J. L., Nishii, L., Leslie, L. M., Lun, J., Lim, B. C., … Aycan, Z. (2011). Differences between tight and loose cultures: A 33-nation study. *Science, 332*(6033), 1100–1104.

Gill, M. J., & Ungson, N. D. (2018). How much blame does he truly deserve? Historicist narratives engender uncertainty about blameworthiness, facilitating motivated cognition in moral judgment. *Journal of Experimental Social Psychology, 77*, 11–23.

Goette, L., Huffman, D., & Meier, S. (2006). The impact of group membership on cooperation and norm enforcement: Evidence using random assignment to real social groups. *American Economic Review, 96*(2), 212–216.

Goldberg, J. H., Lerner, J. S., & Tetlock, P. E. (1999). Rage and reason: The psychology of the intuitive prosecutor. *European Journal of Social Psychology, 29*(5–6), 781–795.

Goldstein, N. J., Cialdini, R. B., & Griskevicius, V. (2008). A room with a viewpoint: Using social norms to motivate environmental conservation in hotels. *Journal of Consumer Research, 35*(3), 472–482.

Goodwin, G. P., Piazza, J., & Rozin, P. (2014). Moral character predominates in person perception and evaluation. *Journal of Personality and Social Psychology, 106*(1), 148.

Graeber, D. (2015). *The utopia of rules: On technology, stupidity, and the secret joys of bureaucracy*. Melville House.

Granot, Y., Balcetis, E., Schneider, K. E., & Tyler, T. R. (2014). Justice is not blind: Visual attention exaggerates effects of group identification on legal punishment. *Journal of Experimental Psychology: General, 143*(6), 2196.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108*(4), 814.

Hamlin, J. K., Wynn, K., Bloom, P., & Mahajan, N. (2011). How infants and toddlers react to antisocial others. *Proceedings of the National Academy of Sciences, 108*, 19931–19936.

Hannikainen, I. R., Tobia, K. P., De Almeida, G. D. F., Donelson, R., Dranseika, V., Kneer, M., … Struchiner, N. (2021). Are there cross-cultural legal principles? Modal reasoning uncovers procedural constraints on law. *Cognitive Science, 45*(8), e13024.

Hart, H. L. A. (1961). *The concept of law*. Oxford University Press.

Henrich, J., Bowles, S., Boyd, R. T., Hopfensitz, A., Richerson, P. J., Sigmund, K., … Young, H. P. (2003). Group report: The cultural and genetic evolution of human cooperation. In P. Hammerstein (Ed.), *Genetic and cultural evolution of cooperation* (pp. 445–468). MIT Press.

Henrich, J., Ensminger, J., McElreath, R., Barr, A., Barrett, C., Bolyanatz, A., … Ziker, J. (2010). Markets, religion, community size, and the evolution of fairness and punishment. *Science, 327*, 1480–1484.

Henrich, J., & Henrich, N. (2006). Culture, evolution and the puzzle of human cooperation. *Cognitive systems research, 7*(2–3), 220–245.

Huebner, B. M., & Bynum, T. S. (2008). The role of race and ethnicity in parole decisions. *Criminology, 46*, 907–938.

Inbar, Y., Pizarro, D. A., & Cushman, F. (2012). Benefiting from misfortune: When harmless actions are judged to be morally blameworthy. *Personality and Social Psychology Bulletin, 38*(1), 52–62.

Jacobs, D., Carmichael, J. T., & Kent, S. L. (2005). Vigilantism, current racial threat, and death sentences. *American Sociological Review, 70*(4), 656–677.

Jensen, K., Call, J., & Tomasello, M. (2007). Chimpanzees are rational maximizers in an ultimatum game. *Science, 318*, 107–109.

Jost, J. T., Glaser, J., Kruglanski, A. W., & Sulloway, F. J. (2003). Political conservatism as motivated social cognition. *Psychological Bulletin, 129*(3), 339.

Kahan, D. M. (2013). Ideology, motivated reasoning, and cognitive reflection: An experimental study. *Judgment and Decision making, 8*, 407–424.

Knobe, J. (2003). Intentional action and side effects in ordinary language. *Analysis, 63*(3), 190–194.

Knobe, J. (2010). Person as scientist, person as moralist. *Behavioral and Brain Sciences, 33*(04), 315–329. https://doi.org/10.1017/S0140525X10000907

Kruglanski, A. W., Shah, J. Y., Fishbach, A., Friedman, R., Chun, W., & Sleeth-Keppler, D. (2002). A theory of goal systems. *Advances in Experimental Social Psychology, 34*.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin, 108*(3), 480.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13).

LaCosse, J., & Quintanilla, V. (2021). Empathy influences the interpretation of whether others have violated everyday indeterminate rules. *Law and Human Behavior, 45*(4), 287.

Landy, J. F., Piazza, J., & Goodwin, G. P. (2016). When it's bad to be friendly and smart: The desirability of sociability and competence depends on morality. *Personality and Social Psychology Bulletin, 42*(9), 1272–1290.

Leach, C. W., Ellemers, N., & Barreto, M. (2007). Group virtue: The importance of morality (vs. competence and sociability) in the positive evaluation of in-groups. *Journal of Personality and Social Psychology, 93*(2), 234.

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2020). Emmeans: Estimated marginal means. CRAN. (R package version 1.4.4).

Levi, M., & Sacks, A. (2009). Legitimating beliefs: Sources and indicators. *Regulation & Governance, 3*(4), 311–333.

Li, J., & Zhou, X. (2014). Sex, attractiveness, and third-party punishment in fairness consideration. *PLoS One, 9*(4), Article e94004.

Lindström, B., Jangard, S., Selbing, I., & Olsson, A. (2018). The role of a "common is moral" heuristic in the stability and change of moral norms. *Journal of Experimental Psychology: General, 147*(2), 228.

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods, 47*(4), 1122–1135.

Malle, B. F., Guglielmo, S., & Monroe, A. E. (2014). A theory of blame. *Psychological Inquiry, 25*(2), 147–186.

Marlowe, F. W., Berbesque, J. C., Barr, A., Barrett, C., Bolyanatz, A., Cardenas, J. C., … Tracer, D. (2008). More 'altruistic' punishment in larger societies. *Proceedings of the Royal Society B: Biological Sciences, 275*(1634), 587–592.

McAuliffe, K., Jordan, J. J., & Warneken, F. (2015). Costly third-party punishment in young children. *Cognition, 134*, 1–10.

McLaughlin, E. C. (2018, August 27). *Why Serena Williams' catsuit ban matters, and what it says about us*. CNN. Retrieved from https://www.cnn.com/2018/08/27/tennis/serena-williams-catsuit-ban-racism-misogyny/index.html.

Mende-Siedlecki, P. & Havlicek, L. (n.d). The Delaware behavior database: A set of social behaviors and corresponding norming data. Norming data and stimuli available at https://osf.io/wg8by/.

Miller, D. T., & Prentice, D. A. (1996). The construction of social norms and standards. In E. T. Higgins, & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 799–829). Guilford Press.

Miller, D. T., & Prentice, D. A. (2016). Changing norms to change behavior. *Annual Review of Psychology, 67*(1), 339–361.

Nadler, J. (2012). Blaming as a social process: The influence of character and moral emotion on blame. *Law and Contemporary Problems, 75*(2), 1–31.

Nadler, J., & McDonnell, M. H. (2011). Moral character, motive, and the psychology of blame. *Cornell Law Review, 97*, 255.

Paharia, N., Vohs, K. D., & Deshpandé, R. (2013). Sweatshop labor is wrong unless the shoes are cute: Cognition can both help and hurt moral motivated reasoning. *Organizational Behavior and Human Decision Processes, 121*(1), 81–88.

Pierson, E., Simoiu, C., Overgoor, J., Corbett-Davies, S., Jenson, D., Shoemaker, A., … Goel, S. (2020). A large-scale analysis of racial disparities in police stops across the United States. *Nature Human Behaviour, 4*, 736–745. https://doi.org/10.1038/s41562-020-0858-1

Pizarro, D. A., & Tannenbaum, D. (2012). *Bringing character back: How the motivation to evaluate character influences judgments of moral blame*.

Qualtrics, I. (2018). *Qualtrics. Provo, UT, USA*.

R Core Team. (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Raggs, T., & Boren, C. (2018, Sep 08). Serena Williams fined $17,000 for U.S. Open outburst; Billie Jean King calls out 'double standard.' The Washington Post. www.washingtonpost.com/news/early-lead/wp/2018/09/08/serena-williamss-game-penalty-at-u-s-open-final-sparks-torrent-of-reactions/.

Raihani, N. J., & McAuliffe, K. (2012). Human punishment is motivated by inequity aversion, not a desire for reciprocity. *Biology Letters, 8*(5), 802–804.

Rattazzi, A. M. M., Bobbio, A., & Canova, L. (2007). A short version of the right-wing authoritarianism (RWA) scale. *Personality and Individual Differences, 43*(5), 1223–1234.

Riedl, K., Jensen, K., Call, J., & Tomasello, M. (2012). No third-party punishment in chimpanzees. *Proceedings of the National Academy of Sciences, 109*, 14824–14829.

Robinson, P. H., & Darley, J. M. (1995). *Justice, liability, and blame: Community views and the criminal law*. Boulder, CO: Westview Press.

Robinson, T. E., & Berridge, K. C. (1993). The neural basis of drug craving: An incentive-sensitization theory of addiction. *Brain Research Reviews, 18*(3), 247–291.

Samoylov, M., Kuntzman, G., Colon, D., & Cuba, J. (2020, January 28). NYPD Targets Blacks and Latinos for 'Jaywalking' Tickets. Retrieved from https://nyc.streetsblog.org/2020/01/08/nypd-targets-blacks-and-latinos-for-jaywalking-tickets/.

Schauer, F. (2015). *The force of law*. Harvard University Press.

Schiller, B., Baumgartner, T., & Knoch, D. (2014). Intergroup bias in third-party punishment stems from both ingroup favoritism and outgroup discrimination. *Evolution and Human Behavior, 35*(3), 169–175.

Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J., & Griskevicius, V. (2007). The constructive, destructive, and reconstructive power of social norms. *Psychological Science, 18*(5), 429–434.

Searle, J. R., & Willis, S. (1995). *The construction of social reality*. Simon and Schuster.

Singmann, H., & Kellen, D. (2019). An introduction to mixed models for experimental psychology. In *New methods in cognitive psychology* (pp. 4–31). Routledge.

Soss, J., Langbein, L., & Metelko, A. R. (2003). Why do white Americans support the death penalty? *The Journal of Politics, 65*(2), 397–421.

Southern Poverty Law Center. (2018, October 18). Alabama's War on Marijuana.. Retrieved from https://www.splcenter.org/20181018/alabamas-war-marijuana?eType=EmailBlastContent&eId=fbfb3108-c152-47d8-8e23-28fac65ab617.

Struchiner, N., Hannikainen, I., & Almeida, G. (2020). An experimental guide to vehicles in the park. *Judgment and Decision Making, 15*(3).

Thompson, B. Y. (2017). "Good moral characters": How drug felons are impacted under state marijuana legalization laws. *Contemporary Justice Review, 20*(2), 211–226.

Tomasello, M. (2009). *Why we cooperate*. MIT press.

Tomasello, M. (2014). The ultra-social animal. *European Journal of Social Psychology, 44*(3), 187–194.

Turri, J. (2019). Excuse validation: A cross-cultural study. *Cognitive Science, 43*(8), Article e12748.
Turri, J., & Blouw, P. (2015). Excuse validation: A study in rule-breaking. *Philosophical Studies, 172*(3), 615–634.
Tyler, T. R. (2006). *Why people obey the law*. Princeton University Press.
Tyler, T. R., & Blader, S. L. (2005). Can businesses effectively regulate employee conduct? The antecedents of rule following in work settings. *Academy of Management Journal, 48*(6), 1143–1158.
Tyler, T. R., Callahan, P. E., & Frost, J. (2007). Armed, and dangerous (?): Motivating rule adherence among agents of social control. *Law and Society Review, 41*(2), 457–492.
Uhlmann, E. L., Pizarro, D. A., Tannenbaum, D., & Ditto, P. H. (2009). The motivated use of moral principles. *Judgment and Decision making, 4*(6), 479–491.
Weber, M. (1921/2002). *Wirtschaft und gesellschaft: Grundriss der verstehenden Soziologie*. Mohr Siebeck.
Weidman, A. C., Sowden, W. J., Berg, M. K., & Kross, E. (2020). Punish or protect? How close relationships shape responses to moral violations. *Personality and Social Psychology Bulletin, 46*(5), 693–708.
Westfall, J. (2016). *PANGEA [computer program]*.
Westfall, J., Kenny, D. A., & Judd, C. M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *Journal of Experimental Psychology: General, 143*(5), 2020.
Whitley, B. E., Jr. (1999). Right-wing authoritarianism, social dominance orientation, and prejudice. *Journal of Personality and Social Psychology, 77*(1), 126.
Wilson, G. D. (Ed.). (1973). *The psychology of conservatism*. London: Academic Press.
Yudkin, D. A., Rothmund, T., Twardawski, M., Thalla, N., & Van Bavel, J. J. (2016). Reflexive intergroup bias in third-party punishment. *Journal of Experimental Psychology: General, 145*(11), 1448.
Zakrisson, I. (2005). Construction of a short version of the right-wing authoritarianism (RWA) scale. *Personality and Individual Differences, 39*(5), 863–872.
Zeisel, H. (1981). Race bias in the administration of the death penalty: The Florida experience. *Harv. L. Rev., 95*, 456.
Zhang, J., Berridge, K. C., Tindell, A. J., Smith, K. S., & Aldridge, J. W. (2009). A neural computational model of incentive salience. *PLoS Computational Biology, 5*(7), Article e1000437.