American Psychological Association
American Psychology-Law Society
Law and Human Behavior

© 2023 American Psychological Association
ISSN: 0147-7307

2023, Vol. 47, No. 2, 367–383
https://doi.org/10.1037/lhb0000527

Moral Appraisals Guide Intuitive Legal Determinations

Brian Flanagan[1], Guilherme F. C. F. de Almeida[2], Noel Struchiner[3], and Ivar R. Hannikainen[4]
[1] School of Law and Criminology, Maynooth University
[2] Law School, Insper Institute of Education and Research
[3] Department of Law, Pontifical Catholic University of Rio de Janeiro
[4] Department of Philosophy I, Faculty of Psychology, University of Granada

Objectives: We sought to understand how basic competencies in moral reasoning influence the application of private, institutional, and legal rules. ***Hypotheses:*** We predicted that moral appraisals, implicating both outcome-based and mental state reasoning, would shape participants' interpretation of rules and statutes—and asked whether these effects arise differentially under intuitive and reflective reasoning conditions. ***Method:*** In six vignette-based experiments (total $N = 2,473$; 293 university law students [67% women; age bracket mode: 18–22 years] and 2,180 online workers [60% women; mean age = 31.9 years]), participants considered a wide range of written rules and laws and determined whether a protagonist had violated the rule in question. We manipulated morally relevant aspects of each incident—including the valence of the rule's purpose (Study 1) and of the outcomes that ensued (Studies 2 and 3), as well as the protagonist's accompanying mental state (Studies 5 and 6). In two studies, we simultaneously varied whether participants decided under time pressure or following a forced delay (Studies 4 and 6). ***Results:*** Moral appraisals of the rule's purpose, the agent's extraneous blameworthiness, and the agent's epistemic state impacted legal determinations and helped to explain participants' departure from rules' literal interpretation. Counter-literal verdicts were stronger under time pressure and were weakened by the opportunity to reflect. ***Conclusions:*** Under intuitive reasoning conditions, legal determinations draw on core competencies in moral cognition, such as outcome-based and mental state reasoning. In turn, cognitive reflection dampens these effects on statutory interpretation, allowing text to play a more influential role.

Public Significance Statement
When deciding whether someone has violated a written rule, people initially consult their moral instincts about the incident. With more time to reflect, their interpretation draws closer to the letter of the law. This finding suggests that in frontline settings, in which time for reflection is scarce (e.g., law enforcement), a rule's application will depend significantly on the interpreter's moral values.

Keywords: textualism, time pressure, moral reasoning, legal interpretation, theory of mind

Supplemental materials: https://doi.org/10.1037/lhb0000527.supp

On July 11, 2019, Roderick Jones, an Australian national, failed in his bid to persuade the High Court of Ireland that he was eligible for naturalized Irish citizenship. Jones, the court ruled, had not satisfied Section 15 of the Irish Nationality and Citizenship Act of 1956, which requires that applicants have had "a period of one year's continuous residence in the State immediately before the date of the application" (*Jones v. Minister for Justice & Equality*, 2019, para. 2) Invoking the Oxford English Dictionary, the court

Bradley D. McAuliff served as Action Editor.

Brian Flanagan https://orcid.org/0000-0001-7719-5051
Guilherme F. C. F. de Almeida https://orcid.org/0000-0002-9134-9843
Noel Struchiner https://orcid.org/0000-0001-6237-9691
Ivar R. Hannikainen https://orcid.org/0000-0003-0623-357X

This research was made possible through Maynooth University's Faculty of Social Sciences and Social Sciences Institute awarded to Brian Flanagan; the Spanish Ministry of Science, Innovation and Universities (PID2020-119791RA-I00) awarded to Ivar R. Hannikainen; the National Council of Scientific and Technological Development of Brazil (309735/2019-0) awarded to Noel Struchiner; the Carlos Chagas Filho Research Support Foundation (E-26/201.071/2021); and the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (001). The authors have no known conflicts of interest to disclose.

Study data, analysis scripts, and materials can be found on Open Science Framework at https://osf.io/gfmcx/. Preregistrations for Studies 2 (https://aspredicted.org/yu28f.pdf), 3 (https://aspredicted.org/ji8ba.pdf), and 4 (https://aspredicted.org/fh7a2.pdf) can be accessed on https://AsPredicted.org.

Brian Flanagan played lead role in funding acquisition and writing of original draft and equal role in conceptualization, investigation, methodology and writing of review and editing. Guilherme F. C. F. de Almeida played lead role in formal analysis and visualization and equal role in conceptualization, investigation, methodology and writing of review and editing. Noel Struchiner played equal role in conceptualization, investigation, methodology and writing of review and editing. Ivar R. Hannikainen played equal role in conceptualization, formal analysis, funding acquisition, investigation, methodology, validation and writing of review and editing.

Correspondence concerning this article should be addressed to Ivar R. Hannikainen, Department of Philosophy I, Faculty of Psychology, University of Granada, Cartuja Campus, 18011 Granada, Spain. Email: ivar@ugr.es

interpreted Section 15 as prescribing an "unbroken" physical presence, which precluded applicants from taking so much as a day trip outside the country during the 12-month period. The court's interpretation dovetailed with a strand of legal scholarship that identifies the application of a rule with the application of its text (Hart, 1958; Schauer, 2009). Jones appealed. Rejecting the High Court's analysis as "overly literal" (para. 47), the Court of Appeal ultimately favored a "purposive, reasonable and pragmatic approach" (para. 59) to Section 15, allowing for up to 6 weeks of foreign travel (*Jones v. Minister for Justice & Equality*, 2019).

Jones is an example of a phenomenon common to jurisdictions in both major legal traditions, common law and civilian: the occasional rejection by judges of the letter of the law in favor of its spirit (MacCormick & Summers, 1991). Numerous studies now confirm that this tendency is not confined to judicial contexts: Laypeople's judgments of whether a rule has been violated are routinely influenced by elements beyond the rule's literal meaning. This effect has been demonstrated in the context of household and institutional rules (Bregant et al., 2019; LaCosse & Quintanilla, 2021; Struchiner et al., 2020), minor legal rules (Garcia et al., 2014; Turri, 2019; Turri & Blouw, 2015), and criminal laws (Kahan, 2010; Peter-Hagene & Bottoms, 2017; Peter-Hagene & Ratliff, 2021).

Purpose or Morality?

In *Jones*, the Court of Appeal does not spell out what exactly led it to dismiss the textually mandated result. This omission reflects the opacity of the phenomenon of spirit-led interpretation (MacCormick & Summers, 1991). Are references to "purpose" meant to convey the provision's historic purpose, that is, the reason why the rule was originally adopted (Alexander & Sherwin, 2008; Goldsworthy, 2005; Sinclair, 1997)? Or is the idea instead that the reasonableness of the agent's conduct should be taken into consideration (Dworkin, 1986; Fuller, 1969; Greenberg, 2014)?

Our first objective in the present research was to dissociate the influence of purpose and morality on rule application (see also de Almeida et al., 2022). Existing research documenting counter-literal determinations (Bregant et al., 2019; Garcia et al., 2014; Struchiner et al., 2020) has focused exclusively on people's reasoning about benevolent rules, that is, rules and laws that were adopted with morally decent purposes in mind. As a result, whether existing evidence of counter-literal decision-making reflects a retrieval of the rule maker's intent or the application of one's prescriptive moral standards is unclear.

A wide literature demonstrates that moral considerations causally influence a broad range of judgments, for example, whether an agent acted intentionally (Kneer & Bourgeois-Gironde, 2017; Knobe, 2003) and whether they caused a negative outcome (Alicke, 2000). For example, we treat someone's past misconduct—but not their good conduct—as causally relevant to their later misfortune, even where there is no plausible link between them (Callan et al., 2014). Similarly, in the context of accidental harm, perpetrators of low moral character are ascribed a greater intention to bring about the harmful outcome than those of high moral standing (Schwartz et al., 2022). In the legal sphere, correlational evidence lends credence to the *moralist* hypothesis that moral appraisals are implicated in legal decision-making. Attitudes of moral condemnation have been found to predict mock trial verdicts (Salerno & Peter-Hagene, 2013; Skorinko et al., 2014) and the application of everyday rules (LaCosse & Quintanilla, 2021). For instance, whether people view an agent's behavior as having violated an applicable rule correlates strongly with their judgments of whether that same behavior is morally blameworthy (Struchiner et al., 2020). Collectively, past research motivates the prediction that morality, not legislative intent, helps to explain people's departure from a strictly textualist application of rules.

Intuition and Reflection in Legal Decision-Making

A further question arises as to the cognitive mechanisms underlying textualist and counter-literal determinations. Which cognitive processes support people's emphasis on the letter and the spirit of rules, respectively? A wealth of past research distinguishes *intuitive* cognitive processes that yield quick and effortless responses from *reflective* processes that demand time and mental effort. For instance, intuitive and reflective processes differentially impact decision-making under risk (Ben Zur & Breznitz, 1981), economic preferences (Teoh et al., 2020), and honesty (Capraro et al., 2019). Establishing how legal determinations depend similarly on the practical features of the decision-making context would help map the connections between the mechanisms of the legal mind and of social cognition in general. Accordingly, our second objective in these studies was to understand whether intuitive and cognitive processes play different roles in supporting deference to a rule's letter or spirit.

Previous evidence motivates both competing hypotheses: on the one hand, comprehensively evaluating an incident's moral status can demand ample cognitive resources (see Kennett & Fine, 2009). For instance, people's moral judgments operate over information about the probability and magnitude of a behavior's morally relevant outcomes (Engelmann & Waldmann, 2022; Shenhav & Greene, 2010) and integrate these representations of outcome value with inferences about the perpetrator's accompanying mental state (Patil & Trémolière, 2021; Young et al., 2010). These results characterize moral reasoning as fairly cognitively demanding, suggesting that counter-literal judgments may be supported by cognitive processing.

On the other hand, numerous studies have documented people's capacity to spontaneously appraise others' behavior (Gigerenzer, 2010; Haidt, 2001). People can morally evaluate others' conduct even under strict time pressure (Tinghög et al., 2016; see also Decety & Cacioppo, 2012), and these initial appraisals are often unaffected by subsequent reflection (McHugh et al., 2017). This body of literature motivates the opposing prediction that counter-literal judgments may be dominant under conditions favoring intuitive judgment.

The Current Research

We conducted six vignette-based experiments administered via a Web browser. These studies describe a series of rules, each with a written formulation and stated purpose. In every study, we manipulated whether the protagonist in the case infringed the text and/or the purpose of the rule, and participants judged whether the protagonist violated the rule. In Studies 1–3, we pursued our first objective: to investigate why people make counter-literal rule determinations. To this end, we manipulated the moral valence of the rule's purpose (Study 1) and the agent's extraneous blameworthiness and character (Studies 2 and 3). We also recorded participants' attitudes of moral blame to ask whether they mediate these experimental effects. In Studies 4–6, we pursued our second objective: to investigate whether textualist and counter-literal decisions are differentially supported by intuitive and cognitive processing. The research was approved by the Maynooth University Research Ethics Committee (SRESC-2020-2411932).

We analyzed the data using mixed-effects linear or logistic regression models in which participants and scenarios were treated as crossed random effects, using the *lmerTest* package (Kuznetsova et al., 2017). To assess the significance of each term in the model, we report F tests with Kenward–Roger degrees of freedom and employ Type 2 sum of squares to facilitate the interpretation of interactions and main effects. In other words, effects (e.g., main effects) are assessed only with other effects of equal or lower order in the model (i.e., leaving out two-way interactions). As a measure of effect size suitable for mixed-effects modeling, we employ the semipartial R^2 (R^2_{sp}) metric introduced by Nakagawa and Schielzeth (2013) and implemented in the *r2glmm* package (Jaeger, 2017). Study data, analysis scripts, and materials are available on Open Science Framework (OSF) at https://osf.io/gfmcx/. An estimation of statistical power indicated that mean post hoc power across the six studies ranged between 82% and 94% (see Table S1 in the online Supplemental Materials).

Study 1: Moral Versus Evil Purposes

Mounting evidence shows that people deviate from a strictly textualist judgment pattern: for instance, if in order to prevent accidents, a nuclear power plant institutes a rule restricting access to the control room (e.g., “Access for trained personnel only”), a research assistant who forgets their training may be seen as violating the rule by entering the control room. Existing research leaves open whether this counter-literal tendency reflects a consideration of the rule maker’s intention (e.g., a deference to their supervisors’ goal of maintaining a safe environment) or an exercise in moral reasoning (e.g., a personal appraisal that the research assistant’s behavior was blameworthy). To discriminate between these explanations, in Study 1, we introduced rules that serve immoral purposes.

Suppose that the “access for trained personnel only” rule was instead introduced to maintain the secrecy of an inhumane research program on the effects of radiation exposure. A human rights activist completes the required training and enters the control room. Did they violate the power plant’s rule? In this second case, the purposivist and moralist theories make distinct predictions: if rule violation judgments reflect a consideration of rule makers’ objectives (i.e., to keep their inhumane research secret), then this case will be seen as a violation. If, by contrast, rule violation judgments are shaped by moral reasoning, then the research assistant with a commitment to human rights will be seen as blameless and, therefore, as having complied with the rule. To further test the moralist prediction, we included a measure of participants’ personal moral condemnation of the incidents, which served as a candidate mediator in subsequent statistical models.

For each rule, we formulated pairs of moral and immoral purposes. The set of immoral purposes was inspired by real-world rule-making vices, including harsh utilitarian calculation, group-based prejudice, and autocratic caprice (see full materials on OSF at https://osf.io/gfmcx/).

Method

Participants

One hundred twenty-seven students (age bracket mode: 18–22 years; 38 men [30%], 80 women [63%], and nine undeclared [7%]) were recruited in an introductory legal course at a large university in Ireland and completed the survey. No incentives were provided for participation.

Materials and Measures

Study 1 employed four scenarios: *power plant access*, *no touching*, *speed limit*, and *firearm control*. In each scenario, a rule-making authority responds to a perceived challenge by adopting a particular text for either a moral or an immoral purpose. In the no-touching scenario, a family seeks either to safeguard its racial purity or to avoid a dangerous viral infection by adopting the rule, “Do not touch anyone outside the home.” In the speed-limit scenario, a king seeks either to amuse himself at the expense of his subjects or to improve road safety by adopting the rule, “It is an offense to drive at any speed in excess of 30 km/h.” And in the firearm-control scenario, a parliament seeks either to suppress an indigenous community’s rite of passage into adulthood or to protect an endangered animal population by adopting the rule, “It is an offense to use any firearm in the vicinity of a wild boar.” An agent was then described who violated both the rule’s text and its purpose (*text-and-purpose* case), violated the text while abiding by the purpose (*text-only* case), violated the purpose while abiding by the text (*purpose-only* case), or violated neither (*neither-text-nor-purpose* case).

For each case, participants made a rule violation judgment and a subjective disapproval judgment on separate 7-point Likert scale. The dependent measure was participants’ level of agreement with a statement that the agent had violated the rule (1 = *Strongly disagree*, 4 = *Neither agree nor disagree*, 7 = *Strongly agree*). The mediator was participants’ judgments of whether the agent had done “a bad thing” (from 1 = *strongly disagree* to 7 = *strongly agree*).

Procedure

The study followed a 2 (text: abide vs. violate) × 2 (purpose: abide vs. violate) × 2 (valence: moral vs. immoral) mixed design in which we randomly assigned participants to either the moral or immoral valence condition, with the remaining variations occurring within subjects. In both the moral and immoral valence conditions, participants were assigned to blocks in which they viewed all four combinations of text and purpose paired with a different scenario in a random order.

For example, in one block, participants viewed the neither-text-nor-purpose case paired with the power-plant-access scenario, the text-only case paired with the speed-limit scenario, the text-and-purpose case paired with the no-touching scenario, and the purpose-only case paired with the firearm-control scenario. Immediately after each case, participants reported whether the agent violated the rule. At the end of the experiment, participants were asked whether the agent in each case had done “a bad thing.”

Hypotheses and Planned Analyses

The purposivist hypothesis (Hypothesis 1) was that we would observe a main effect of purpose and no Purpose × Valence interaction—such that participants would report higher violation judgments when a rule’s purpose is violated than when it is not (regardless of its moral valence). In contrast, the moralist hypothesis (Hypothesis 2) was that we would observe a Purpose × Valence interaction—such that participants would report higher violation judgments when the purpose is violated in the moral than the immoral valence condition. To test these predictions, we regressed violation judgments in a mixed-effects linear regression model with text, purpose, moral valence, and every two- and three-way interaction as fixed effects.

Results

We report descriptive statistics of rule violation judgments by condition in Table 1. The model ($R^2_m = .44$) revealed main effects of text, $F(1, 360) = 287.12$, $R^2_{\mathrm{sp}} = .13$; purpose, $F(1, 350) = 56.47$, $R^2_{\mathrm{sp}} = .005$; and moral valence, $F(1, 125) = 21.49$, $R^2_{\mathrm{sp}} = .001$, all ps $< .001$. Critically, we observed a two-way Purpose × Valence interaction, $F(1, 362) = 29.38$, $R^2_{\mathrm{sp}} = .03$, $p < .001$. No other terms achieved statistical significance. Examining the marginal effect of purpose violation separately for moral and immoral purposes yielded support for the moralist hypothesis (Hypothesis 2): violating a morally good purpose promoted rule violation judgments, $B = 2.02$, $t = 9.21$, $p < .001$, whereas violating an immoral purpose did not, $B = 0.35$, $t = 1.62$, $p = .11$ (see Figure 1).

Participants' disapproval ratings correlated with their rule violation judgments, $r_{\mathrm{partial}}(484) = .60$, $p < .001$—when we partialed out the variance attributable to scenario (replicating the results of Struchiner et al., 2020). To assess whether the experimental effects observed in our primary model were explained by participants' personal attitudes of disapproval, we entered disapproval ratings as an additional predictor in a second mixed-effects model. In this model ($R^2_m = .58$), disapproval of the agents' behavior predicted rule violation judgments, $F(1, 435) = 138.91$, $R^2_{\mathrm{sp}} = .009$, $p < .001$. The main effects of text, $F(1, 372) = 215.20$, $R^2_{\mathrm{sp}} = .064$, $p < .001$, and purpose, $F(1, 389) = 7.21$, $R^2_{\mathrm{sp}} = .006$, $p = .008$, remained significant, whereas the main effect of valence, $F(1, 202) = 0.42$, $R^2_{\mathrm{sp}} = .002$, $p = .52$, and the Purpose × Valence interaction, $F(1, 448) = 0.54$, $R^2_{\mathrm{sp}} = .001$, $p = .46$, did not.

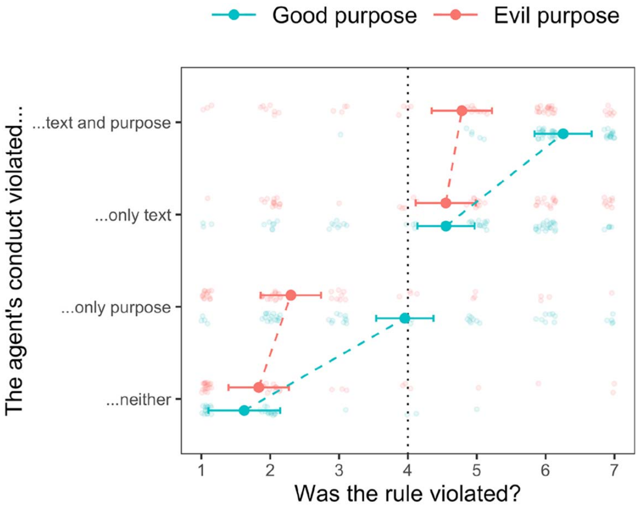
The results of the second model—showing that personal disapproval absorbed the variance associated with the Purpose × Valence interaction—suggest that the selective effect of moral-purpose violations may be mediated by participants' personal disapproval of the agents' conduct. To investigate this relationship, we conducted a mediation analysis (with the *mediation* package; Tingley et al., 2014) and found that the Purpose × Valence interaction effect on rule violation judgments (i.e., the selective effect of purpose violation when the purpose is morally good) was mediated by subjective moral disapproval, $B = 1.93$, $p < .001$, rendering the direct effect nonsignificant, $B = –0.27$, $p = .42$.

Table 1
Study 1 Descriptive Statistics

Purpose	Text	Valence	*n*	*M*	*SD*	95% CI *LL*	95% CI *UL*
Not violated	Not violated	Bad	60	1.83	1.45	1.47	2.20
Not violated	Not violated	Good	45	1.62	0.91	1.36	1.89
Not violated	Violated	Bad	60	4.55	2.05	4.03	5.07
Not violated	Violated	Good	67	4.55	1.87	4.10	5.00
Violated	Not violated	Bad	60	2.30	1.55	1.91	2.69
Violated	Not violated	Good	67	3.96	2.14	3.44	4.47
Violated	Violated	Bad	60	4.78	1.93	4.29	5.27
Violated	Violated	Good	67	6.25	1.20	5.97	6.54

Note. CI = confidence interval; *LL* = lower limit; *UL* = upper limit.

Figure 1
Study 1: Means and 95% Confidence Intervals for Each Factorial Combination of Text, Purpose, and Valence



Note. The dashed lines illustrate the simple effects of purpose (good vs. evil) on rule application. See the online article for the color version of this figure.

Discussion

In line with prior research, our results showed that a rule's text exerted a dominant influence on violation judgments (Struchiner et al., 2020). Yet simultaneously manipulating the moral valence of the rule's purpose and whether the purpose was undermined by the target act uncovered evidence for the moralist hypothesis (Hypothesis 2): agents who undermined the purpose of a benevolent rule were more likely to be judged as violating the rule than those who did not, whereas no such effect arose for agents who undermined the purpose of an evil rule. Furthermore, this effect was mediated by participants' personal disapproval of the agent's behavior. In sum, rule makers' historical intentions mattered only when they were morally good—revealing that counter-literal verdicts are primarily the product of moral evaluation and not of deference to a rule's purpose per se.

Study 2: Good Versus Evil Agents

Study 1 provided preliminary evidence that morality affects rule violation judgments. Our focus was on the manipulation of the moral valence of the rule's purpose, yet participants may have justifiably inferred differences in the agent's moral character across conditions as well. Thus, whether counter-literal judgments were driven solely by moral evaluations of the rule's purpose (as suggested by Hart & Sacks, 1994) or by a more comprehensive appraisal of the incident (as observed by LaCosse & Quintanilla, 2021; Wylie & Gantman, 2023) remains unclear. To shed light on the question, in Study 2, we held the purposes' moral valence constant, describing a benevolent rule in every case, and extraneously manipulated the protagonist's moral blameworthiness.

Method

Participants

One hundred sixty-eight students (age bracket mode: 18–22 years; 56 men [33%] and 112 women [67%]) recruited in introductory courses at a large university in Ireland completed the survey. No incentives were provided for participation.

Materials and Measures

We employed four scenarios involving benevolent rules: *no dogs*, *only native citizens can run for president*, *no sleeping on the station benches*, and *government subvention for home ownership*. For instance, in the no-dogs scenario, participants first read the following introductory paragraph:

> In Penfold City, citizens and tourists frequently complain that their experience of eating in a restaurant can be ruined by the unruly behavior of other diners' pet dogs. To meet these concerns, city officials decided to introduce a new ordinance: "No dogs allowed in the restaurants of Penfold City".

We then described an agent who had violated either the text (text-only case; e.g., a blind person enters the restaurant with their well-behaved guide dog) or the purpose (purpose-only case; e.g., a performer enters a local restaurant with a misbehaving pet monkey) of the provision. We introduced an orthogonal manipulation of agent *blameworthiness*: in the high-blame/text-only condition, the pet owner not only violated the rule's text but also was guilty of animal cruelty (carrying their dog in an uncomfortable carrier after a surgical operation), whereas in the low-blame/text-only condition, the pet owner was suitably caring. The dependent measure was rule violation judgments recorded on a 7-point Likert scale, as in Study 1.

Procedure

In a 2 (blameworthiness: high vs. low) × 2 (case type: text only vs. purpose only) between-subjects design, participants evaluated four cases presented in a random order. After each case, participants judged whether the agent had broken the rule.

Hypotheses and Planned Analyses

We predicted (Hypothesis 1) a main effect of agent blameworthiness—specifically, that judgments of whether the rule had been violated would be higher in the high blameworthiness condition than the low blameworthiness condition (see https://aspredicted.org/yu28f.pdf for the preregistration). The preregistered test of our hypothesis was a mixed-effects linear model of rule violation judgments with case type, blameworthiness, and the Case Type × Blameworthiness interaction as fixed effects.

Results

We report descriptive statistics of rule violation judgments by condition in Table 2. Our preregistered model ($R^2_m = .11$) revealed a main effect of blameworthiness, $F(1, 452) = 41.70$, $R^2_{\text{sp}} = .10$, and an interaction between blameworthiness and case type, $F(1, 165) = 20.66$, $R^2_{\text{sp}} = .03$, both ps $< .001$, but no main effect of case type, $F(1, 165) = 0.10$, $R^2_{\text{sp}} = .02$, $p = .75$. As illustrated in Figure 2, the simple effect of blameworthiness attained statistical significance in text-only cases, $B = 1.77$, $t = 7.69$, $p < .001$, but not purpose-only cases, $B = 0.37$, $t = 1.78$, $p = .080$.

Table 2
Study 2 Descriptive Statistics

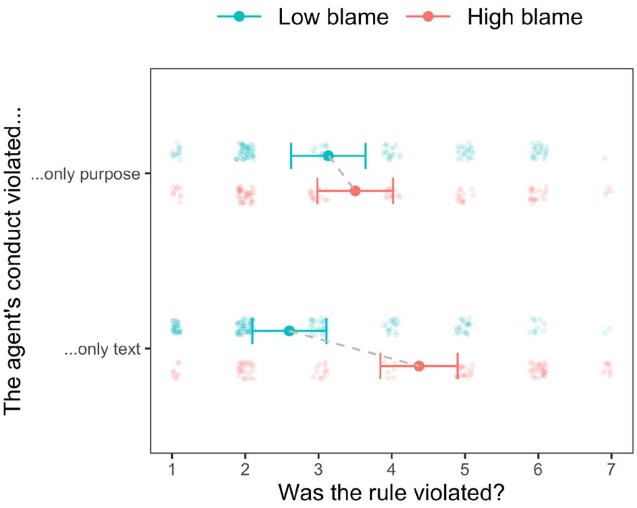
					95% CI	
Case type	Blameworthiness	*n*	*M*	*SD*	*LL*	*UL*
Text only	High	140	4.37	2.13	4.02	4.72
Text only	Low	216	2.60	1.60	2.39	2.82
Purpose only	High	141	3.49	1.97	3.16	3.81
Purpose only	Low	175	3.14	1.79	2.88	3.41

Note. CI = confidence interval; *LL* = lower limit; *UL* = upper limit.

Discussion

An extraneous manipulation of the agent's moral blameworthiness impacted participants' determinations of whether the corresponding rule had been violated. This effect emerged even when we held the rule's text and purpose constant, suggesting that the role of moral reasoning in rule application is not circumscribed to a consideration of the rule's purpose. Rather, it encompasses various morally relevant aspects of the incident in question, including the agent's character and the outcomes that ensue. Unexpectedly, the effect of moral blameworthiness on rule violation judgments arose only for text-only cases. Perhaps an agent's blamelessness plays an exculpatory role when, from a purely textualist perspective, they

Figure 2
Study 2: Means and 95% Confidence Intervals for Each Factorial Combination of Case Type and Blameworthiness



Note. The dashed lines illustrate the simple effects of blameworthiness (low vs. high) on rule application. See the online article for the color version of this figure.

have violated the rule—whereas their blameworthiness does not play an inculpating role when they have abided by the rule's text. In Study 3, we pursued this explanation further.

Study 3: Distinguishing Morality and Legality

Studies 1 and 2 documented an impact of moral appraisals (of both the rule's purpose and extraneous moral considerations) on legal determinations. However, participants decided the cases through a single rule violation item—which raises concerns about an artifactual explanation. In particular, affirming that an action violates a rule could carry the implication that the perpetrator deserves punishment or sanction (Turri & Blouw, 2015). If so, in certain circumstances, participants might believe that an action constituted a rule violation but report that it did not, for example, if they wanted to convey that the agent should be exempt from punishment.

In Study 3, participants had the opportunity to simultaneously and independently report whether the target acts were immoral and whether they violated the rule in question. If the previously observed effect of moral blameworthiness on participants' rule application was an artifact of our previous experimental design, we would expect no effect to arise when participants are able to concurrently manifest their moral condemnation using separate items. Conversely, if the effect persists in these circumstances, we could conclude that moral evaluation is a genuine influence on rule violation.

Finally, in line with "inclusive" legal positivism, legal systems might be presumed to have constitutional constraints in place requiring that legal outcomes be morally acceptable (e.g., Hart, 1979). To preclude this effect on participants' decisions, we stipulated that the fictional country in which these scenarios take place had no such constitutional criteria (e.g. "Penfold City is an independent city state, whose constitution assigns unfettered legislative power to an elected assembly and omits any mention of individual rights.").

Method

Instrument Development

We first drafted a list of 16 items: eight statements that the agent had violated the rule and eight statements that their conduct or attitudes were immoral. We recruited 75 participants from https://Prolific.co to report their level of agreement with each statement on a 7-point Likert scale. The statements were presented in a random order across participants. An exploratory factor analysis confirmed that the scale was composed of two factors: a rule violation factor and a blameworthiness factor (see Table 3). We then selected three items per factor on the basis of face validity and factor loadings while ensuring that we retained one reverse-coded item in each factor. Both three-item measures demonstrated excellent internal validity (rule violation: Cronbach's α = .93; moral blameworthiness: Cronbach's α = .86).

Participants

Participants were 254 https://Prolific.co workers who were paid £0.75 each. Forty-one participants failed a preregistered attention check and were excluded from subsequent analyses. Our final sample comprised 213 participants (M_{age} = 27 years; 84 men [39%], 128 women [60%], and one nonbinary [<1%]).

Materials and Measures

We employed the same four scenarios as in Study 2. Each scenario had an introductory paragraph describing an incident that gave rise to the adoption of a written rule. Next, the vignettes narrated a case involving a violation of both the rule's text and its purpose (text-and-purpose case), the text (text-only case), the purpose (purpose-only case), or neither (neither-text-nor-purpose case), carried out by either a blameworthy (high blame) or a neutral (low blame) agent. Altogether, this resulted in eight experimental

Table 3
Study 3 Factor Loadings and Uniqueness

Item	Factor loading		Uniqueness
	1	2	
[Agent] is disobeying the legislature.	0.93		0.12
[Agent] is not complying with the legislation.[a]	0.90		0.17
[Agent] violated the statute.[a]	0.89		0.19
[Agent] is breaking the rules.[a]	0.88		0.20
[Agent] is ignoring the law.	0.85		0.25
[Agent] is playing by the rules. (reverse-scored)	–0.84		0.26
[Agent] is abiding by the rules. (reverse-scored)	–0.82		0.32
[Agent] is entitled to [perform act]. (reverse-scored)	–0.73	–0.30	0.38
What [agent] would do is out of line.		0.94	0.11
What [agent] would do is wrong.[a]		0.93	0.10
[Agent] deserves to be reprimanded for [performing act].[a]		0.80	0.34
I would rather [agent] not [perform act].		0.73	0.43
What [agent] would do is blameworthy.	0.30	0.74	0.36
In my opinion, [agent] is a good person. (reverse-scored)[a]		–0.72	0.48
[Agent] should not get to [perform act].	0.46	0.55	0.48
It is OK for [agent] to [perform act]. (reverse-scored)	–0.49	–0.59	0.41

Note. The content in brackets varied across scenarios and provided case-specific detail. We dropped the negation from the "complying with the legislation" item to form our reverse-scored item in Study 3.
[a] This item was retained in Study 3.

conditions per scenario. Table S2 in the online Supplemental Materials presents the verbatim text of the no-dogs scenario. The dependent variable was the three-item average of rule violation (Cronbach's α = .90), and the mediator variable was the three-item average of moral blame (Cronbach's α = .88).

Procedure

In a 2 (text: abide vs. violate) × 2 (purpose: abide vs. violate) × 2 (blameworthiness: high vs. low) balanced incomplete block design, participants were randomly assigned to eight blocks and viewed four scenarios in each block. Each of the four case types (text-and-purpose, text-only, purpose-only, neither-text-nor-purpose) was paired with a different scenario on each trial. In each block, blameworthiness was counterbalanced, such that two scenarios involved a blameworthy agent and two scenarios involved a neutral agent. For each case, participants answered the rule violation and moral blame items in a randomized order.

Hypotheses and Planned Analyses

The artifact hypothesis (Hypothesis 1) was that we would observe no effects of blameworthiness after accounting for the main effects of text, purpose, and the Text × Purpose interaction. In contrast, the moralist hypothesis (Hypothesis 2a) was that blameworthiness would influence rule violation judgments after we accounted for the main effects of text, purpose, and the Text × Purpose interaction. In light of the asymmetry between purpose-only and text-only cases in Study 2, we preregistered a specific version of the moralist hypothesis, in which (Hypothesis 2b) the positive effect of moral blame would be stronger when the rule's text was violated (i.e., a Blameworthiness × Text interaction; see https://aspredicted.org/ji8ba.pdf for the preregistration).

The first preregistered model regressed violation judgments (i.e., the three-item average) on text, purpose, blameworthiness, and every two- and three-way interaction as fixed effects. A second preregistered model drew on the three-item moral blame measure and included the following variables as fixed effects: the text, purpose, and blameworthiness factors; the moral blame measure; every interaction between text, purpose, and blameworthiness; and every interaction between text, purpose, and the moral blame measure. The purpose of this second analysis was to ascertain whether the effect of blameworthiness in the first model would be accounted for by participants' attitudes of moral blame.

Results

We report descriptive statistics of rule violation judgments by condition in Table 4. Our preregistered model (R^2_m = .25) showed significant main effects of text, $F(1, 630) = 285.37$, $R^2_{sp} = .051$, and purpose, $F(1, 630) = 33.56$, $R^2_{sp} = .008$, both $ps < .001$. In support of the moralist hypothesis (Hypothesis 2a), we observed a main effect of blameworthiness as well, $F(1, 630) = 7.44$, $R^2_{sp} = .002$, $p = .006$. Contrary to our specific prediction (Hypothesis 2b), results showed that the Blameworthiness × Text interaction was not significant, $F(1, 211) = 0.69$, $R^2_{sp} = .001$, $p = .79$. None of the remaining terms achieved statistical significance, $ps > .10$. The simple effect of blameworthiness was significant in text-only cases, $B = –0.66$, $t = –2.92$, $p = .003$, but nonsignificant in the remaining cases, $–0.32 < Bs < –0.01$, $|t|s < 1.42$, $ps > .15$ (see Figure 3).

Table 4
Study 3 Descriptive Statistics

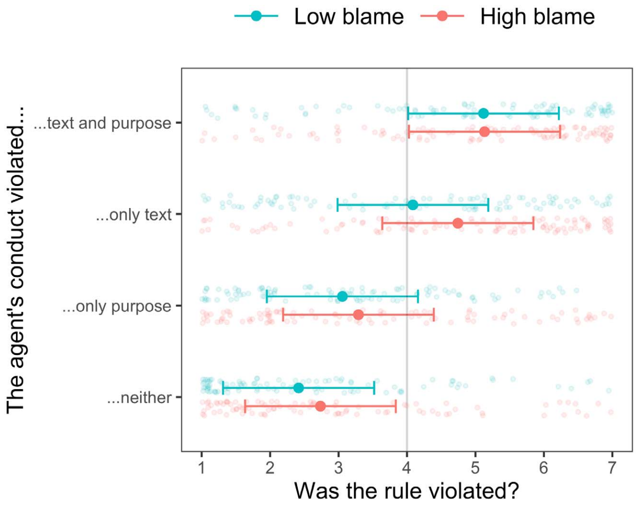
Text	Purpose	Blame	*n*	*M*	*SD*	95% CI	
						LL	*UL*
Not violated	Not violated	Low	110	2.41	1.59	2.12	2.71
Not violated	Not violated	High	103	2.76	1.73	2.43	3.10
Not violated	Violated	Low	110	3.06	1.62	2.76	3.36
Not violated	Violated	High	103	3.29	1.89	2.93	3.66
Violated	Not violated	Low	103	4.06	1.91	3.69	4.43
Violated	Not violated	High	110	4.74	1.83	4.40	5.08
Violated	Violated	Low	103	5.11	1.77	4.77	5.45
Violated	Violated	High	110	5.14	1.69	4.82	5.45

Note. CI = confidence interval; *LL* = lower limit; *UL* = upper limit.

Our second model ($R^2_m = .27$) uncovered significant main effects of text, $F(1, 629) = 316.05$, $R^2_{sp} = .038$; purpose, $F(1, 642) = 20.03$, $R^2_{sp} = .013$; and subjective judgments of the agent's moral blame, $F(1, 837) = 28.94$, $R^2_{sp} = .020$, all $ps < .001$. Mirroring Study 1, results showed that including the moral blame measure in the model rendered the effect of our experimental manipulation of blameworthiness nonsignificant, $F(1, 746) = 2.20$, $R^2_{sp} = .004$, $p = .14$. Once again, we did not observe the predicted Moral Blame × Text interaction, $F(1, 805) = 0.17$, $R^2_{sp} = .001$, $p = .68$.

The model, however, revealed an unpredicted Moral Blame × Purpose interaction, $F(1, 838) = 3.83$, $R^2_{sp} = .008$, $p = .050$, which was qualified by a three-way interaction with text, $F(1, 807) = 4.40$, $R^2_{sp} = .004$, $p = .036$. This result reflects the weaker association between moral blame and rule violation judgments in purpose-only cases, $B = 0.03$, $p = .74$, than in any other condition, $Bs > 0.21$, $ps < .005$.

Figure 3
Study 3: Means and 95% Confidence Intervals for Each Factorial Combination of Text, Purpose, and Blameworthiness



Note. See the online article for the color version of this figure.

To further probe the relationship between the blameworthiness manipulation and participants' ratings of moral blame, we conducted a mediation analysis. Results conceptually replicated Study 1, showing that blameworthiness indirectly affected rule violation judgments via attitudes of moral blame, $B = 0.38$, $p < .001$, rendering the direct effect nonsignificant, $B = –0.06$, $p = .70$.

Discussion

Moral blameworthiness influenced legal determinations even while participants concurrently reported their moral attitudes toward the agent—ruling out the possibility that the effect was an artifact of our single-item measure. Contrary to our preregistered prediction, the impact of blameworthiness did not depend on whether the text had been violated. Rather, the effect was weakest in purpose-only cases—a pattern that also held in Study 2, pointing toward a disjunctive explanation of the role of moral appraisals: Blameworthy conduct—whether in violation of a benign purpose or otherwise—may independently suffice to elicit a counter-literal judgment, whereas the presence of both blameworthy elements does not elevate counter-literal responses above the extent to which either element does so on its own.

Study 4: Effects of Time Pressure on Legal Determinations

Studies 1–3 provided convergent evidence that moral appraisals guide people's determinations of whether rules have been violated. This effect arises whether agents violate the purpose of a benign rule (Study 1) or provide extraneous reasons to be considered morally blameworthy (Studies 2 and 3). What cognitive processes might support the counter-literal judgment pattern documented throughout these studies? In Study 4, we employed a manipulation of *time pressure* (e.g., Suter & Hertwig, 2011), randomly assigning participants to decide cases either within a few seconds or after a forced delay, to investigate whether intuition and reflection play dissociable roles in legal decision-making.

On the one hand, evidence that judging an incident's moral status may demand cognitive resources (see Kennett & Fine, 2009; Shenhav & Greene, 2010) motivates the *reflective-moralism* hypothesis: Reflection time allows participants to engage in a comprehensive moral evaluation that results in counter-literal judgments, whereas legal determinations align more closely with literal meaning under time pressure. On the other hand, the *intuitive-moralism* hypothesis coheres with evidence that people routinely issue quick moral judgments (Gigerenzer, 2010; Haidt, 2001). If so, legal determinations under time pressure ought to reflect a stronger influence on moral attitudes. Comparatively, a forced delay may enable participants to reach a textualist verdict, for example, if cognitive control is required to override an intuitive moral preference. A further possibility is that intuitive and reflective processes support textualist and counter-literal judgments equally—in which case manipulating time pressure would uncover no overall effect.

Method

Participants

Participants were 450 https://Prolific.co workers who were paid £0.61 each. Ninety-one participants failed a preregistered attention check and were excluded from subsequent analyses. Our final sample comprised 359 participants (mean age = 35 years; 130 men [36%], 224 women [62%], five nonbinary [1%]).

Materials and Measures

We employed eight brief scenarios involving benevolent rules: *no dogs*, *no cars in the park*, *no sleeping on the benches*, *no smartphones in class*, *no shoes in the apartment*, *no touching allowed*, *no shooting*, and *only authorized personnel*. The scenarios were shortened as an adaptation to the time pressure paradigm and were composed of three statements: (a) a background statement (e.g., "A friend entered Mary's house wearing shoes and dirtied the carpets"), (b) a rule-introduction statement (e.g., "To keep her house clean, Mary announced: 'No one may wear shoes in the house'"), and (c) a conduct statement, which varied by case type.

To increase the generalizability of our results in Study 4, we randomly displayed one of three conduct statements within each factorial combination of text, purpose, and scenario (as shown in Table S3 in the online Supplemental Materials). As a further adaptation of the experimental protocol to the speeded response paradigm, we reduced the number of scale points in the dependent measure: in Study 4, participants provided dichotomous responses to the rule violation question ("Did this person violate the rule?"; 1 = *yes*, 0 = *no*).

Procedure

In a 2 (condition: speeded vs. delayed) between-subjects × 2 (text: violated vs. not violated) × 2 (purpose: violated vs. not violated) within-subjects design, each participant was randomly assigned to either the speeded or delayed condition. In each condition, participants assessed four different cases corresponding to the text and purpose within-subjects manipulations—the order of which was randomized across participants. For each case, the statements were presented on separate screens and displayed for 5 s each. On the final screen, participants recorded a rule violation judgment.

The novel, between-subjects manipulation was the time participants were given to provide a response. In the speeded condition, participants had to answer within 4 s. Meanwhile, in the delayed condition, participants were unable to submit their responses until 15 s had elapsed. Additionally, to foster reflection in the delayed condition, the background, introduction, and conduct statements were reintroduced alongside the rule violation question, and participants were instructed to briefly explain their answers in writing.

Hypotheses and Planned Analyses

Both competing hypotheses imply two-way interaction effects of condition with the text and/or purpose factors. The intuitive-textualism hypothesis would predict (Hypothesis 1a) a positive Condition × Text interaction, reflecting a stronger effect of text under time pressure than after a forced delay and/or (Hypothesis 1b) a negative Condition × Purpose interaction, reflecting a weaker effect of purpose under time pressure than after a forced delay. The intuitive-moralism hypothesis would predict (Hypothesis 2a) a negative Condition × Text interaction, reflecting a weaker effect of text under time pressure than after a forced delay and/or (Hypothesis 2b) a positive Condition × Purpose interaction, reflecting a stronger effect of purpose under time pressure than after a forced delay. We did not preregister any directional hypothesis (see https://aspredicted.org/fh7a2.pdf for the preregistration).

The preregistered model regressed rule violation judgments on condition, text, purpose, and the Condition × Text and Condition × Purpose interactions in a mixed-effects logistic model. We stipulated that we would run the primary analysis twice: first excluding only inattentive participants (i.e., who failed the attention check) and, second, additionally excluding noncompliant participants (i.e., who took longer than 4 s to decide in the speeded condition or wrote gibberish in the justification question in the delayed condition). In combination, these analyses tested whether there was a causal effect (in the *intent-to-treat* analysis) that was not driven by noncompliance (in the *treatment-on-the-treated* analysis).

Results

We report descriptive statistics of rule violation judgments by condition in Table 5.

Intent-to-Treat Analysis

A likelihood-ratio test of the preregistered model (ΔR^2_m = .50) revealed main effects of text, $\chi^2(1)$ = 220.15, R^2_{sp} = .108, and purpose, $\chi^2(1)$ = 121.27, R^2_{sp} = .048, both *ps* < .001, but not condition, $\chi^2(1)$ = 0.01, R^2_{sp} = .001, *p* = .93. We observed a Condition × Text interaction, $\chi^2(1)$ = 11.40, R^2_{sp} = .005, *p* < .001, but no Condition × Purpose interaction, $\chi^2(1)$ = 1.43, R^2_{sp} = .001, *p* = .23. In line with the intuitive-moralism hypothesis (Hypothesis 2a), the Condition × Text interaction revealed that the marginal effect of text was weaker in the speeded condition, odds ratio (*OR*) = 22.32, *z* = 12.44, than in the delayed condition, *OR* = 67.57, *z* = 12.85, both *ps* < .001.

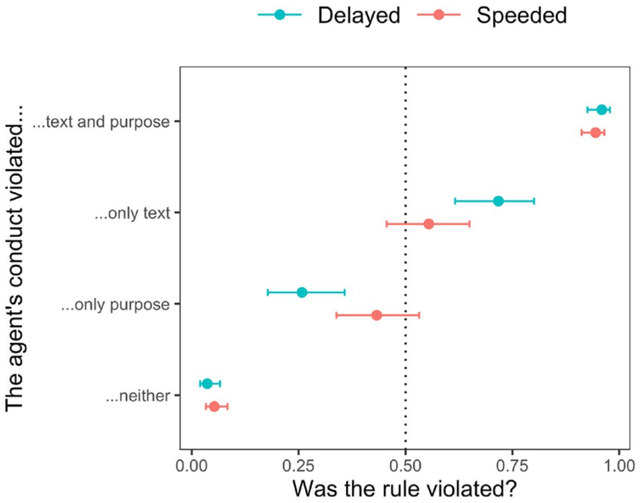
Examination of the simple effects of condition for each case type revealed no influence of time pressure when textualist and moral interpretation supported the same verdict, that is, in text-and-purpose cases, *z* = 1.00, *p* = .32, and neither-text-nor-purpose cases, *z* = –1.23, *p* = .22. Meanwhile, for conflict cases (characterized by opposing textual and moral interpretations), time pressure impacted people's responses: with additional time to reflect, participants were more likely to deem text-only cases violations, *z* = 3.66, and less likely to view purpose-only cases as violations, *z* = –3.76, than when judging under time pressure, both *ps* < .001 (see Figure 4).

Table 5
Study 4 Descriptive Statistics

Text	Purpose	Time pressure	*n*	$p_{(\text{Yes})}$	95% CI *LL*	95% CI *UL*
Not violated	Not violated	Delayed	187	.06	.03	.10
Not violated	Not violated	Speeded	344	.07	.05	.11
Not violated	Violated	Delayed	187	.26	.20	.33
Not violated	Violated	Speeded	344	.43	.38	.49
Violated	Not violated	Delayed	187	.67	.60	.74
Violated	Not violated	Speeded	344	.52	.47	.58
Violated	Violated	Delayed	187	.96	.92	.98
Violated	Violated	Speeded	344	.94	.90	.96

Note. CI = confidence interval; *LL* = lower limit; *UL* = upper limit.

Figure 4
Study 4: Probability and 95% Confidence Intervals for Each Factorial Combination of Text, Purpose, and Time Pressure



Note. See the online article for the color version of this figure.

Treatment-on-the-Treated Analysis

Our second analysis (ΔR^2_m = .53) revealed convergent results: when excluding noncompliant participants, we observed significant main effects of text, $\chi^2(1) = 211.01$, $R^2_{\text{sp}} = .104$, and purpose, $\chi^2(1) = 113.71$, $R^2_{\text{sp}} = .030$, both *ps* < .001, but not condition, $\chi^2(1) = 0.27$, $R^2_{\text{sp}} = .00$, *p* = .61. Once again, the Condition × Text interaction was significant, $\chi^2(1)$ = 6.09, R^2_{sp} = .003, *p* = .014, whereas the Condition × Purpose interaction was not, $\chi^2(1)$ = 3.30, R^2_{sp} = .002, *p* = .069. The pattern of simple effects remained the same as in our primary model: time pressure had no effect on either text-and-purpose cases, *z* = 0.39, *p* = .70, or neither-text-nor-purpose cases, *z* = –0.28, *p* = .78, but strengthened counter-literal tendencies in both text-only cases, *z* = 3.77, and purpose-only cases, *z* = –3.31, both *ps* < .001.

Discussion

Under time pressure, participants were more likely to report counter-literal verdicts in text-only and purpose-only cases, providing support for the intuitive-moralism hypothesis. Reflective processes amplified the impact of text on rule application, strengthening the tendency toward textualist determinations when participants had unlimited time to reason. In this way, Study 4 documented the dissociable roles of intuitive and reflective cognitive processes in legal decision-making, as previously observed in other domains of social cognition (e.g., Callan et al., 2010).

Study 5: Knowledge Versus Foreseeability

Our first four studies examined the effects of moralization arising from a consideration of outcome value: the manipulation of purpose violation contrasted cases in which either a bad or a neutral outcome ensues (e.g., when the floors get dirty vs. stay clean in the context of a "no shoes in the house" rule). It is known, however, that moral judgments are also influenced by the agent's mental state (Almagro et al., 2022; Kirfel & Phillips, 2023). Offenses that are carried out intentionally elicit stronger disapproval than those same offenses brought about accidentally (Young et al., 2010)—a distinction that is reflected in the law's definition of mens rea as a formal element of serious crime (Gardner, 1993). Accordingly, the application of rules may depend on agents' foreknowledge of the rules' purpose. For instance, dirtying the floor might constitute a clearer violation of the "no shoes in the house" rule if the guest is aware that the rule's purpose is to keep the floors clean. This example illustrates how foreknowledge (relative to ignorance) of a rule's purpose may promote counter-literal verdicts.

Study 5 simultaneously examined an orthogonal prediction about what the agent should have known (Kirfel & Hannikainen, 2023). When determining a person's degree of culpability for accidental and improbable harms, research has found that people consult their standards about what they ought to have known (Kneer & Skoczeń, 2023; Nobes & Martin, 2022). Relatedly, legal theorists have noted the unfairness of demanding that people abide by a rule's unstated purpose in situations in which this purpose is difficult to discern (Evans, 1988; Sinnott-Armstrong, 2005). Third, in common law systems, determining what an agent ought to have foreseen can result in the ascription of liability for negligence via analogical legal reasoning (e.g., *Donoghue v. Stevenson*, 1932; *United States v. Carroll Towing Co.*, 1947). This view predicts that legal determinations in Study 5 may depend on whether a rule's purpose can be readily inferred from its text—with foreseeable purposes encouraging greater counter-literal determinations than unforeseeable purposes.

Method

Participants

Participants were 419 https://Prolific.co workers who were paid £0.63 each. Fifteen participants failed a preregistered attention check and were excluded from subsequent analyses. Our final sample comprised 404 participants (mean age = 31 years; 196 men [49%], 199 women [49%], nine nonbinary [2%]).

Materials and Measures

We employed five scenarios (no shoes, no cars in the park, no smartphones in class, no dogs, and no shooting), each of which was composed of three statements: (a) a rule-introduction statement (e.g., "The headmaster announced, 'Phones may not be used in the classroom'"); (b) a mental-state statement, which varied depending on the knowledge and foreseeability conditions (e.g., "Jill [knows/does not know] that this rule is in place to [get students to pay attention in class/reduce accessory envy]"); and (c) a conduct statement, which varied by case type (e.g., "Jill texts her friends in class on her brand new Apple watch").

The dependent variable in this study was a rule violation judgment (e.g., "Did Jill violate the rule?") recorded on a 4-point Likert scale (1 = *No*, 2 = *Probably Not*, 3 = *Probably*, 4 = *Yes*).

Procedure

In a 2 (knowledge: present vs. absent) × 2 (foreseeability: high vs. low) × 2 (case type: text only vs. purpose only) balanced incomplete block design, participants were assigned to four between-subjects conditions in the Knowledge × Foreseeability matrix (i.e., present-high, present-low, absent-high, and absent-low) and decided two purpose-only and two text-only cases paired with four different scenarios in a random order.

All vignettes began with a description of the rule's text (e.g., "Mary announced: 'No one may wear shoes in my apartment'"). Depending on condition assignment, the rule's purpose was either foreseeable (e.g., to keep the floors clean) or unforeseeable (e.g., to reduce noise), and the agent either knew or did not know the rule's purpose (e.g., "Sara knows that this rule is intended to keep Mary's floor clean" or "Sara doesn't know that this rule is intended to keep Mary's floor clean"). We included a filler trial in the middle (i.e., third position), which was a text-and-purpose or a neither-text-nor-purpose case—helping to conceal the study hypothesis and allowing for exploratory analyses.

Pretest

We recruited 78 participants to pretest our manipulation of the purposes' foreseeability. Each participant judged five rules in a random order. For each rule (e.g., the "no shoes in the house" rule), participants rated the likelihood of both purposes (to keep the floors clean and to reduce noise) on percentage scales. In a mixed-effects model, we regressed these likelihood ratings on the dichotomous manipulation of foreseeability (with scenario and participant as random effects). As expected, participants viewed the foreseeable purposes (77.5, 95% confidence interval [CI] [66.1, 88.8]) as significantly easier to infer than the unforeseeable purposes (33.6, 95% CI [22.6, 45.0]), $B = 43.9$, $t = 21.52$, $p < .001$.

Hypotheses and Planned Analyses

The knowledge hypothesis (Hypothesis 1) would predict a Knowledge × Case Type interaction, such that case type would exert a larger influence when the agent lacks (vs. has) knowledge of the rule's purpose. The foreseeability hypothesis (Hypothesis 2) would predict a Foreseeability × Case Type interaction, such that case type would exert a larger influence when the rule's purpose is harder (vs. easier) to infer from its text. To test these hypotheses, we regressed rule violation judgments on case type, knowledge, foreseeability, and every two- and three-way interaction.

Results

Descriptive statistics can be found in Table 6. Our preregistered model ($R^2_m = .05$) revealed main effects of case type, $F(1, 1{,}206) = 9.61$, $R^2_{\text{sp}} = .024$, $p = .002$, and foreseeability, $F(1, 400) = 19.48$, $R^2_{\text{sp}} = .026$, $p < .001$, but not knowledge, $F(1, 400) = 0.03$, $R^2_{\text{sp}} = .001$, $p = .85$. Furthermore, the effect of case type was moderated by foreseeability, $F(1, 1{,}204) = 57.27$, $R^2_{\text{sp}} = .022$, $p < .001$, but not knowledge, $F(1, 400) = 0.12$, $R^2_{\text{sp}} = .001$, $p = .73$.

Decomposing the two-way interaction provided support for the foreseeability hypothesis (Hypothesis 2): when the rule's purpose was unforeseeable, text-only cases were seen as violating the rule significantly more than purpose-only cases (see Figure 5). Furthermore, the marginal effects of case type indicated that participants

Table 6
Study 5 Descriptive Statistics

Knowledge	Foreseeability	Case type	*n*	*M*	*SD*	95% CI *LL*	95% CI *UL*
No	Low	Text only	192	3.34	1.02	3.20	3.49
No	Low	Purpose only	192	2.57	1.34	2.38	2.76
No	High	Text only	208	2.54	1.23	2.38	2.71
No	High	Purpose only	208	2.78	1.28	2.61	2.96
Yes	Low	Text only	192	3.21	1.13	3.05	3.37
Yes	Low	Purpose only	192	2.66	1.19	2.49	2.82
Yes	High	Text only	216	2.59	1.27	2.42	2.76
Yes	High	Purpose only	216	2.80	1.24	2.64	2.97

Note. CI = confidence interval; *LL* = lower limit; *UL* = upper limit.

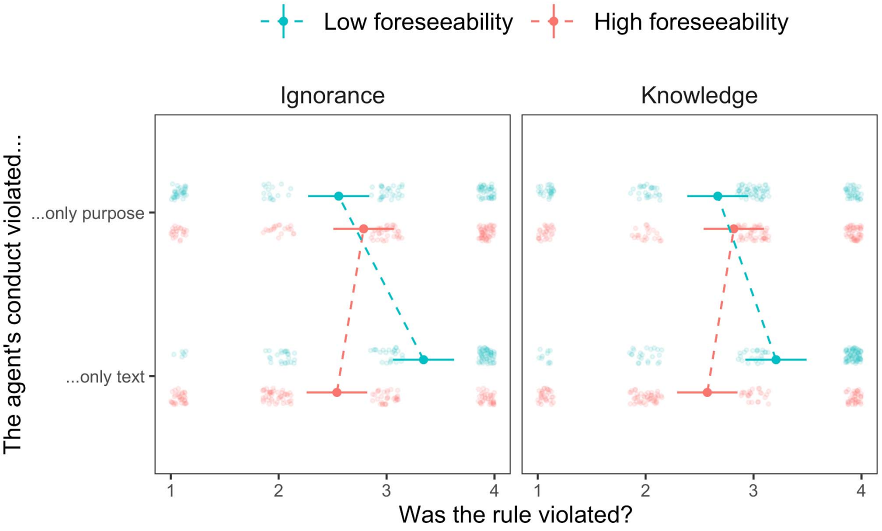
made textualist judgments regardless of whether the agent had knowledge, $B = 0.54$, $t = 4.39$, or lacked knowledge, $B = 0.79$, $t = 6.38$, of the rule's obscure purpose, both $ps < .001$. The difference between case types reversed when the purpose was foreseeable: participants judged that purpose-only cases violated the rule to a greater extent than text-only cases. Once again, this pattern was observed whether we stipulated that the agent knew, $B = –0.25$, $t = –2.11$, $p = .035$, or did not know, $B = –0.25$, $t = –2.10$, $p = .036$, the rule's obvious purpose (see Figure 5).

Including data from the control conditions (i.e., the filler trial), we replaced the case-type term in our primary model with the text (abide vs. violate) and purpose (abide vs. violate) dummy codes. This analysis confirmed that the tendency toward counter-literal determinations in the context of foreseeable purposes was due to a greater impact of the purpose—Foreseeability × Purpose: $F(1, 1981) = 31.99$, $p < .001$; Foreseeability × Text: $F(1, 1982) = 2.64$, $p = .10$. Nonconformity with a foreseeable purpose ($B = 1.52$) constituted a clearer rule violation than nonconformity with an unforeseeable purpose ($B = 0.82$, $t = 5.66$, $p < .001$), whereas nonconformity with text played a comparable role regardless of the purpose's foreseeability, $t = –1.62$, $p = .10$.

Discussion

Overall, participants appeared to disregard agents' actual knowledge of a rule's purpose when deciding whether they had violated the rule. In contrast, participants' knowledge standards played a prominent role: participants placed a weak emphasis on whether unforeseeable purposes had been violated—treating purpose-only cases as compliant with the rule and text-only cases as violations. Comparatively, participants placed a greater emphasis on whether foreseeable purposes had been violated: When a rule's purpose was foreseeable, the tendency toward textualist judgments disappeared and in fact reversed. These patterns arose regardless of whether we stipulated that the agent knew or did not know the rule's purpose—and despite ample time to reflect.

Figure 5
Study 5: Means and 95% Confidence Intervals for Each Factorial Combination of Case Type, Knowledge, and Foreseeability



Note. The dashed lines illustrate the simple effects of case type on rule application. See the online article for the color version of this figure.

Study 6: Foreseeability Under Time Pressure

In Study 4, moral considerations played a greater role under time pressure than after a forced delay. The goal of Study 6 was to examine the complementary hypothesis that the influence of agents' mental states would differ in the comparison between intuitive and reflective reasoning conditions. One possibility is that reasoning about what the agent should have known and integrating this knowledge standard into one's final legal determination demands time and effortful mental state reasoning. If so, the effect of knowledge standards—operationalized as the discrepancy between rules with foreseeable and unforeseeable purposes—ought to arise primarily in the delayed condition.

Other research attests to people's intuitive capacity for mental state ascription (e.g., Decety & Cacioppo, 2012)—as reflected in our intuitive evaluations of a particular agent's mental state against the standard of what an average agent (Tobia, 2018), or perhaps we ourselves (Epley et al., 2004), would believe in the circumstances described. This account makes a different prediction, namely that the observed effect of foreseeability stems from intuitive processes and, therefore, arises even under time pressure. To discriminate between these competing hypotheses, we expanded on the design of Study 5 by manipulating whether participants decided under time pressure or after a forced delay.

Method

Participants

Participants were 1,263 https://Prolific.co workers who were compensated with £0.60 each for approximately 3 min of their time. After we excluded 59 participants who failed our attention check, our final sample was made up of 1,204 participants (mean age = 32 years; 440 men [37%], 755 women [63%], nine nonbinary [1%]).

Materials and Measures

We employed the same five scenarios as in Study 5. Each scenario was composed of three statements: (a) a rule-introduction statement (as in Study 5); (b) a mental-state statement, which varied depending on the knowledge and foreseeability conditions (e.g., "[Jill knows that this/This] rule is in place to [get students to pay attention in class/reduce accessory envy]"); and (c) a conduct statement, which varied by case type (as in Study 5). Rule-violation judgments (e.g., "Did Jill break the rule?") were recorded on a 4-point Likert scale as in Study 5.

Procedure

The procedure was an extension of Study 5, with two primary changes. First, we added a between-subjects manipulation of time pressure (as in Study 4): each case was composed of three statements displayed consecutively for 5 s each before the response slide. Presentation of the response slide, with the rule violation question, varied in the speeded versus delayed conditions just as in Study 4.

For the sake of completeness, we retained the manipulation of the agent's (actual) knowledge state. Unlike Study 5, Study 6 compared the ascription of knowledge not to the ascription of ignorance but merely to the absence of any ascription. This decision was guided by the concern that representing a state of ignorance would be too demanding when participants were deciding under time pressure.

Hypotheses and Planned Analyses

The reflective-mentalizing hypothesis (Hypothesis 1) would predict a negative Condition × Case Type × Foreseeability interaction—such that the negative effect of foreseeability on textualist determinations would be stronger after a forced delay than under time pressure. The intuitive-mentalizing hypothesis would predict (Hypothesis 2a) a two-way Case Type × Foreseeability interaction or even (Hypothesis 2b) a positive Condition × Case Type × Foreseeability interaction—such that the negative effect of foreseeability on textualist determinations would be either comparable in magnitude or weaker after a forced delay than under time pressure.

In our primary analysis, we entered rule violation judgments in a regression model with case type, time pressure, knowledge state, and knowledge standard as fixed effects. The four terms in the model were allowed to interact with each other, except that the knowledge state and knowledge standards factors—which were conceived as candidate moderators of the case type, time pressure, and Case Type × Time Pressure effects—were not allowed to interact with each other.

Results

We report descriptive statistics of rule violation judgments by condition in Table 7. Our primary model (R^2_m = .51) revealed significant main effects of case type, $F(1, 4{,}278) = 109.88$, R^2_{sp} = .016, and knowledge standard, $F(1, 850) = 59.83$, R^2_{sp} = .009, both ps < .001. Case type interacted with time pressure (replicating Study 4), $F(1, 4{,}277) = 37.11$, R^2_{sp} = .001. and with knowledge standard (replicating Study 5), $F(1, 4{,}278) = 67.80$, R^2_{sp} = .003, both ps < .001. The Case Type × Knowledge interaction was again nonsignificant, $F(1, 4{,}278) = 0.19$, R^2_{sp} = .00, p = .66—and there were no other effects of knowledge, all R^2_{sp}s < .001, ps > .30.

In addition, Study 6 uncovered a Condition × Case Type × Foreseeability interaction, $F(1, 4{,}278) = 4.31$, R^2_{sp} = .002, p = .038. This three-way interaction lent support to the intuitive-mentalizing hypothesis (Hypothesis 2b) that foreseeability spontaneously informs rule violation judgments. Specifically, the Case Type × Foreseeability interaction was already present in the speeded condition, $F(1, 2{,}222) = 52.04$, R^2_{sp} = .019, p < .001: Text-only cases were seen as more compliant with the rule when its purpose was readily foreseeable (vs. unforeseeable), B = –0.59, t = –8.98, p < .001. A nonsignificant trend in the opposite direction was found for purpose-only cases, B = 0.09, t = 1.22, p = .18. Meanwhile, in the delayed condition, the Case Type × Foreseeability interaction was weaker, $F(1, 1925) = 19.52$, R^2_{sp} = .007, p < .001—as were the marginal effects of foreseeability by case type (text only: B = –0.48, t = –7.20, p < .001; purpose only: B = –0.06, t = –0.96, p = .34; see Figure 6).

In our closing analysis, we included participants' responses on the filler trial and replaced the case-type term in our primary model with the dichotomous text (abide vs. violate) and purpose (abide vs. violate) factors (R^2_m = .23). In summary, the effect of purpose violation arose under time pressure, was larger for foreseeable than unforeseeable purposes (Purpose × Foreseeability: B = 0.44, t = 6.20, p < .001) and persisted in the delayed condition (Purpose × Condition: B = –0.13,

Table 7
Study 6 Descriptive Statistics

							95% CI	
Knowledge	Foreseeability	Condition	Case type	*n*	*M*	*SD*	*LL*	*UL*
No	Low	Delayed	Text only	302	2.26	1.13	2.13	2.39
No	Low	Delayed	Purpose only	302	1.49	1.25	1.35	1.63
No	Low	Speeded	Text only	294	2.13	1.17	2.00	2.27
No	Low	Speeded	Purpose only	294	1.70	1.22	1.56	1.84
No	High	Delayed	Text only	358	1.81	1.17	1.69	1.93
No	High	Delayed	Purpose only	358	1.49	1.24	1.36	1.62
No	High	Speeded	Text only	350	1.68	1.16	1.56	1.80
No	High	Speeded	Purpose only	350	1.71	1.26	1.58	1.84
Yes	Low	Delayed	Text only	298	2.35	1.06	2.23	2.47
Yes	Low	Delayed	Purpose only	298	1.59	1.27	1.45	1.74
Yes	Low	Speeded	Text only	304	2.25	1.11	2.13	2.38
Yes	Low	Speeded	Purpose only	304	1.69	1.24	1.55	1.83
Yes	High	Delayed	Text only	276	1.87	1.15	1.73	2.01
Yes	High	Delayed	Purpose only	276	1.43	1.24	1.28	1.58
Yes	High	Speeded	Text only	342	1.50	1.22	1.37	1.63
Yes	High	Speeded	Purpose only	342	1.85	1.28	1.72	1.99

Note. CI = confidence interval; *LL* = lower limit; *UL* = upper limit.

$t = -1.90$, $p = .068$). Meanwhile, the effect of text violation also arose under time pressure, regardless of the purpose's foreseeability (Text × Foreseeability: $B = -0.11$, $t = 1.57$, $p = .12$) and was strengthened by the opportunity to reflect (Text × Condition: $B = 0.27$, $t = 3.88$, $p < .001$).

Discussion

Replicating Studies 4 and 5, Study 6 showed that textualist determinations were strengthened by the opportunity to reflect as well as by the difficulty of inferring a rule's purpose from its text. Additionally, the results of Study 6 spoke in favor of the intuitive-mentalizing hypothesis: under time pressure, participants spontaneously considered whether a rule's purpose would be easily inferred from its text and ascribed greater weight to the violation of foreseeable than unforeseeable purposes. This resulted in textualist resolutions when participants interpreted rules with unforeseeable purposes but counter-literal resolutions when they interpreted rules with foreseeable purposes. The effect of knowledge standards on rule violation judgments, if anything, appeared to weaken under reflective conditions—indicating that it did not depend on the availability of cognitive resources.

Taken in conjunction, Studies 4–6 reveal a broader pattern: legal judgments integrate various morally relevant cues, including the agent's epistemic state and the outcomes of their behavior—and these effects arise already in people's intuitive determinations (i.e., under time pressure). The opportunity to reflect appears to strengthen the effect of literal meaning on rule application (see also Hannikainen et al., 2022)—resulting in a shift toward textualist determinations over time.

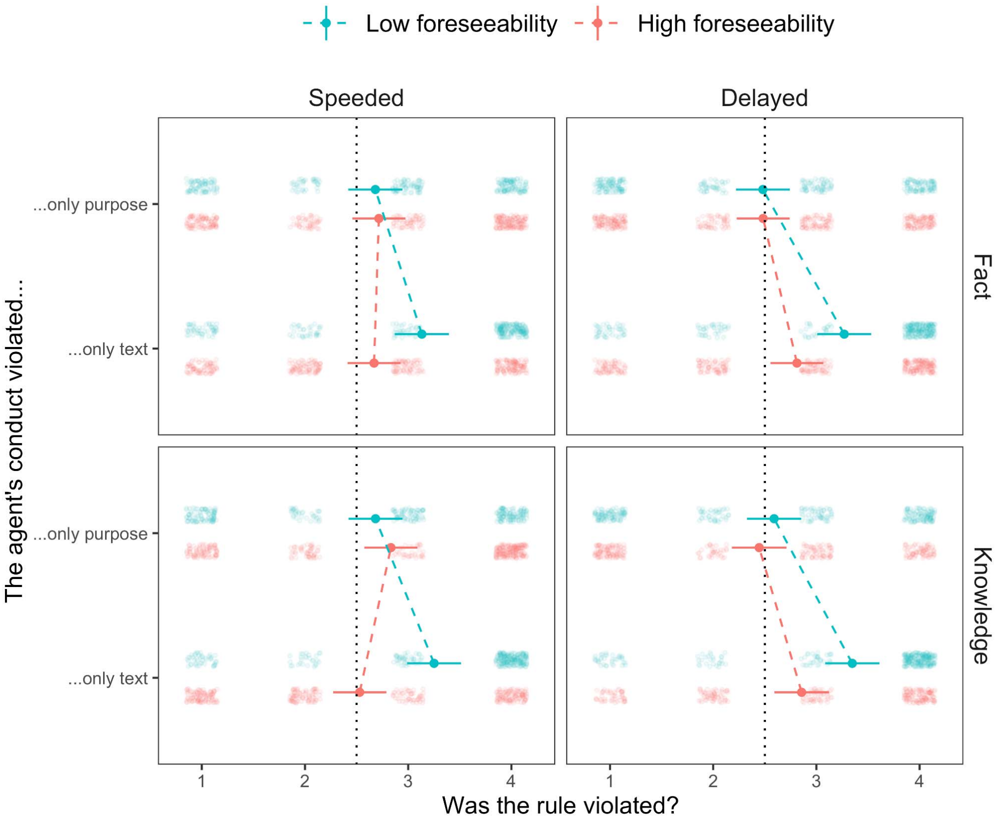
General Discussion

Counter-literal violation judgments emerge when people apply benevolent rules but not rules adopted for evil purposes (Study 1). Similarly, rule application is influenced by extraneous variation in the transgressor's moral blameworthiness (Study 2). These effects persisted when we applied two further robustness checks: (a) when encouraging participants to concurrently and independently evaluate the morality as well as the legality of the target behaviors and (b) when explicitly denying any constitutional constraints on the morality of law (Study 3). Participants' counter-literal decisions also reflected an epistemic standard of what the agent should have known regardless of what they actually knew (Study 5). Turning our attention to the underlying cognitive mechanism, we found evidence that legal decision-making differs in the comparison between the intuitive and reflective reasoning conditions. When deciding under time pressure, participants were more likely to report counter-literal verdicts (Studies 4 and 6). Taken together, our studies provide insight into the origin of counter-literal legal determinations: counter-literal determinations reflect the influence of moral standards and are not merely a retrieval of the rule maker's intent. Moreover, these judgments are significantly more frequent when reached spontaneously and when the rule maker's intent is foreseeable.

These results carry several important implications. At his Senate confirmation hearing, Chief Justice John Roberts of the U.S. Supreme Court described legal interpretation as a matter of calling "balls and strikes" (Roberts, 2005, p. 56). The present finding that rule application is influenced by moral appraisals challenges this view, as well as the broader positivist theory of law that characterizes law and morality as conceptually distinct domains (e.g., Hart, 1958; Raz, 1999; Shapiro, 2011). The challenge posed to legal positivism extends to empirical research programs that adopt a positivist framework. Consider research into the phenomenon of jury nullification (i.e., of trials in which jurors refuse to apply the law and acquit the accused notwithstanding their legal guilt, thereby nullifying the law; Scheflin, 1972, p. 169). Such research identifies jury nullification as occurring when jurors defy the letter of the law (Peter-Hagene & Bottoms, 2017). Accordingly, jurors are said to face a dilemma between "follow[ing] the law, or … engaging in jury nullification … if [doing so] would violate their sense of justice" (Peter-Hagene & Bottoms, 2017, p. 983). Our results suggest that in fact this may sometimes be a false dilemma: the question of whether to find someone guilty of a crime whose actions were morally

Figure 6

Study 6: Means and 95% Confidence Intervals for Each Case Type and Knowledge Condition



Note. The dashed lines illustrate the simple effects of case type on rule application. See the online article for the color version of this figure.

justified may be intuitively internal rather than external to the process of applying the law. The posited tension between jury nullification and the normative ideal of the rule of law (e.g., Horowitz & Kerr, 2001) may therefore be overstated.

Equally, morality's influence on legal determinations has implications for the comparison of legal reasoning with the numerical calculations involved in business accounting (Galligan, 1986; Sterling & Moore, 1987). Theorists have contrasted rule-like norms whose application turns solely on the satisfaction of textually specified conditions with nonmechanical "standards" (Nance, 2006; Pound, 1959; Schlag, 1985). On the basis of this distinction, the choice between enacting rules or standards has been considered a fulcrum through which a variety of conflicting outlooks might shape social outcomes, including individualistic versus altruistic moral worldviews (e.g., Kennedy, 1976), conflicting conceptions of the judicial role (e.g., Sullivan, 1992), alternative understandings of the function of constitutional rights (e.g., Schauer, 2005), and preferences for plutocratic versus egalitarian economic structures (e.g., McBarnet & Whelan, 1991). The current evidence suggests, however, that this analysis may rest on a mistaken analogy; unlike numerical calculation, the interpretation of rules relies on moral appraisals. The thought that rule makers might choose to adopt legal norms that will be applied mechanically might therefore exaggerate their potential influence on social outcomes. In contrast, concern about the social impact of the act of rule application itself has been a theme of empirical research on the role of the law's letter and spirit, especially in relation to policing (Garcia et al., 2014, p. 489; LaCosse & Quintanilla, 2021, p. 305).

A growing literature has documented the influence of bias, most notably racial prejudice, on judicial (Abrams et al., 2012; Cohen & Yang, 2019; Gazal-Ayal & Sulitzeanu-Kenan, 2010) and nonjudicial (Okonofua & Eberhardt, 2015; Weitzer & Brunson, 2015) decision-making. Such bias is exacerbated in frontline contexts, such as law enforcement (Richardson & Goff, 2012), where decision time is limited—and sometimes rectified after the fact in court settings. As noted, our studies demonstrated that textualist interpretation demands time and cognitive resources. This finding may help explain how inconsistent application of a law's text might be driven by certain naturalistic conditions favoring intuition—potentially by conditions characteristic of certain types of policing. In view of the role of the psychological processes that give rise to cognitive and motivational biases in our evaluation of conduct's morality (Alicke, 2000), an important practical question for future research is whether, in intuitive reasoning conditions, rule application draws on broader evaluations, such as stereotypes and biases, as it does on outcome-based and mental state reasoning. Such evidence could help explain why prejudice persists in frontline settings even though it is retrospectively contested in disciplinary or judicial proceedings—where officials have time to reflect on the law's literal scope.

Limitations and Directions for Future Research

The present results relied on a series of contrastive vignettes, drawing certain experimental comparisons of interest through brief, text-based manipulations. As in previous moral psychology research, this approach provides the opportunity to elucidate the factors that causally influence decision-making (Cushman & Greene, 2012)—in particular, the role of the letter versus the spirit—with greater clarity than observational research on real-world behavior allows. This approach helps to consolidate the conclusions drawn from traditional doctrinal analyses of leading cases (Ellinger & Keith, 1999) and qualitative empirical inquiries into the jurisprudential regimes of appellate courts (Gillman, 2001). However, a notable limitation of the contrastive vignette methodology is the question of its ecological validity. The artificiality of our stimulus set may result in low stakes for participants, an overreliance on unrealistic scenarios, and the lack of evidential uncertainty that otherwise pervades real-world decision-making—which together jeopardize the external validity of our results.

A second limitation relates to the generalizability of our findings. Our studies relied on student samples recruited at a large university in Ireland and native English speakers drawn from Prolific, a popular crowdsourcing website. Thus, our results stem from studies conducted exclusively in English—a feature that has recently been diagnosed with the potential to misguide theorizing in cognitive science (Blasi et al., 2022). Still, convenience samples recruited on Prolific have been found to reproduce effects on judgment and decision-making previously observed in studies involving nationally representative samples (Peer et al., 2017). Relatedly, the impact of moral reasoning on rule application documented in the present work has been observed in lay samples throughout numerous countries and in various languages (Hannikainen et al., 2022). Together, these limitations call for further research to investigate whether our current findings arise, for example, in nonindustrialized and small-scale societies and in reaction to more realistic stimuli (e.g., that incorporate a broader range of cultural cues).

Our research indicated that counter-literal judgments were more frequent both when a rule's purpose was morally good and when it was foreseeable. In light of previous research documenting an intuitive tendency to view immoral events as improbable (Phillips & Cushman, 2017; see also Donelson & Hannikainen, 2020), this raises the possibility that well-intentioned rules attract counter-literal applications because it is seen as more likely that rule makers would pursue benevolent over nonbenevolent purposes. Alternatively, rules with foreseeable purposes might attract counter-literal applications because such purposes are perceived to be more morally important. The latter scenario is consistent with the emphasis of moralist legal philosophy on the varying relative weight of text and moral value across alternative questions of legal interpretation (e.g., Dworkin, 1986). Equally, the possibility that some rules have purposes whose perceived moral importance results in more frequent counter-literal application coheres with empirical research describing the impetus, when one is deciding who caused a harmful outcome, to tarnish an agent's conduct as a signal to others of the agent's objectionable tendencies (Knobe & Fraser, 2008; Nadler, 2012). On extending this logic to rule application, it would follow that the greater the moral infraction involved, the greater the potential need to use rule violation judgments to send an equivalent signal or reminder. In future studies, researchers ought to investigate both of these alternative connections.

Conclusion

Formal rules organize large parts of modern society, playing a critical role in community life as well as in institutional and legal settings (Lewis & Steinmo, 2012). If rules' impact "could be assumed independently of the words, the words would be of no use, and the laws of course would not be written" (Spooner, 1847, p. 222). But how people reason about and apply rules is still poorly understood. Although a rule's literal meaning serves as a primary guide to its application, previous research has demonstrated that people often deviate from straightforwardly textualist interpretation. Our present studies help to explain why: people's application of rules is shaped by a spontaneous appraisal of various morally relevant cues—including what the agent ought to have known and the outcomes that ensued. This moral appraisal guides legal determinations most strongly under time pressure; and yet its influence, though attenuated, persists under conditions favoring cognitive control. The role of moral appraisal in rule application underscores a practical limit to the control that authorities can aim to exert on the policing of behavior through their choice of text, namely that "the rule maker cannot … create good judgment where none exists" (Black, 1995, p. 113). People's moral appraisals are not just predictive of their compliance with the law (Gur & Jackson, 2020); they also contribute to their judgments of what it is to be compliant.

References

Abrams, D. S., Bertrand, M., & Mullainathanm, S. (2012). Do judges vary in their treatment of race? *The Journal of Legal Studies*, *41*(2), 347–383. https://doi.org/10.1086/666006

Alexander, L., & Sherwin, E. (2008). *Demystifying legal reasoning*. Cambridge University Press. https://doi.org/10.1017/CBO9781139167420

Alicke, M. D. (2000). Culpable control and the psychology of blame. *Psychological Bulletin*, *126*(4), 556–574. https://doi.org/10.1037/0033-2909.126.4.556

Almagro, M., Hannikainen, I. R., & Villanueva, N. (2022). Whose words hurt? Contextual determinants of offensive speech. *Personality and Social Psychology Bulletin*, *48*(6), 937–953. https://doi.org/10.1177/01461672211026128

Ben Zur, H., & Breznitz, S. J. (1981). The effect of time pressure on risky choice behavior. *Acta Psychologica*, *47*(2), 89–104. https://doi.org/10.1016/0001-6918(81)90001-9

Black, J. M. (1995). "Which arrow?": Rule type and regulatory policy. Public Law, 94–*117*.

Blasi, D. E., Henrich, J., Adamou, E., Kemmerer, D., & Majid, A. (2022). Over-reliance on English hinders cognitive science. *Trends in Cognitive Sciences*, *26*(12), 1153–1170. https://doi.org/10.1016/j.tics.2022.09.015

Bregant, J., Wellbery, I., & Shaw, A. (2019). Crime but not punishment? Children are more lenient toward rule-breaking when the "spirit of the law" is unbroken. *Journal of Experimental Child Psychology*, *178*, 266–282. https://doi.org/10.1016/j.jecp.2018.09.019

Callan, M. J., Sutton, R. M., & Dovalec, C. (2010). When deserving translates into causing: The effect of cognitive load on immanent justice reasoning. *Journal of Experimental Social Psychology*, *46*(6), 1097–1100. https://doi.org/10.1016/j.jesp.2010.05.024

Callan, M. J., Sutton, R. M., Harvey, A. J., & Dawtry, R. J. (2014). Immanent justice reasoning: Theory, research, and current directions. In J. M. Olson & M. P. Zanna (Eds.), *Advances in experimental social psychology*

(Vol. 49, pp. 105–161). Academic Press. https://doi.org/10.1016/B978-0-12-800052-6.00002-0

Capraro, V., Schulz, J., & Rand, D. G. (2019). Time pressure and honesty in a deception game. *Journal of Behavioral and Experimental Economics*, *79*, 93–99. https://doi.org/10.1016/j.socec.2019.01.007

Cohen, A., & Yang, C. S. (2019). Judicial politics and sentencing decisions. *American Economic Journal. Economic Policy*, *11*(1), 160–191. https://doi.org/10.1257/pol.20170329

Cushman, F., & Greene, J. D. (2012). Finding faults: How moral dilemmas illuminate cognitive structure. *Social Neuroscience*, *7*(3), 269–279. https://doi.org/10.1080/17470919.2011.614000

de Almeida, G. F. C. F., Knobe, J., Struchiner, N., & Hannikainen, I. R. (2022). Purposes in law and in life: An experimental investigation of purpose attribution. *Canadian Journal of Law and Jurisprudence*. Advance online publication. https://doi.org/10.1017/cjlj.2022.20

Decety, J., & Cacioppo, S. (2012). The speed of morality: A high-density electrical neuroimaging study. *Journal of Neurophysiology*, *108*(11), 3068–3072. https://doi.org/10.1152/jn.00473.2012

Donelson, R., & Hannikainen, I. R. (2020). Fuller and the folk: The inner morality of law revisited. In T. Lombrozo, J. Knobe, & S. Nichols (Eds.), *Oxford studies in experimental philosophy* (Vol. 3, pp. 6–28). Oxford University Press. https://doi.org/10.1093/oso/9780198852407.003.0002

Donoghue v. Stevenson, UKHL 100 (United Kingdom House of Lords 1932). https://www.lawteacher.net/cases/donoghue-v-stevenson.php

Dworkin, R. (1986). *Law's empire*. Harvard University Press.

Ellinger, E. P., & Keith, K. J. (1999). Legal research: Techniques and ideas. *Law Review*, *30*(2), 459–482. https://doi.org/10.26686/vuwlr.v30i2.6000

Engelmann, N., & Waldmann, M. R. (2022). How to weigh lives. A computational model of moral judgment in multiple-outcome structures. *Cognition*, *218*, Article 104910. https://doi.org/10.1016/j.cognition.2021.104910

Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004). Perspective taking as egocentric anchoring and adjustment. *Journal of Personality and Social Psychology*, *87*(3), 327–339. https://doi.org/10.1037/0022-3514.87.3.327

Evans, J. (1988). *Statutory interpretation: Problems of communication*. Oxford University Press.

Fuller, L. (1969). *The morality of law* (Rev. ed.). Yale University Press.

Galligan, D. J. (1986). *Discretionary powers: A legal study of official discretion*. Oxford University Press.

Garcia, S. M., Chen, P., & Gordon, M. (2014). The letter versus the spirit of the law: A lay perspective on culpability. *Judgment and Decision Making*, *9*(5), 479–490. https://doi.org/10.1017/S1930297500006835

Gardner, M. (1993). The mens rea enigma: Observations on the role of motive in the criminal law past and present. *Utah Law Review*, 635–750.

Gazal-Ayal, O., & Sulitzeanu-Kenan, R. (2010). Let my people go: Ethnic in-group bias in judicial decisions—Evidence from a randomized natural experiment. *Journal of Empirical Legal Studies*, *7*(3), 403–428. https://doi.org/10.1111/j.1740-1461.2010.01183.x

Gigerenzer, G. (2010). Moral satisficing: Rethinking moral behavior as bounded rationality. *Topics in Cognitive Science*, *2*(3), 528–554. https://doi.org/10.1111/j.1756-8765.2010.01094.x

Gillman, H. (2001). What's law got to do with it? Judicial behavioralists test the "legal model" of judicial decision making. *Law & Social Inquiry*, *26*(2), 465–504. https://doi.org/10.1111/j.1747-4469.2001.tb00185.x

Goldsworthy, J. (2005). Legislative intentions, legislative supremacy, and legal positivism. *The San Diego Law Review*, *42*(2), 493–518.

Greenberg, M. (2014). The moral impact theory of law. *The Yale Law Journal*, *123*(5), 1288–1342. https://www.yalelawjournal.org/pdf/1288.Greenberg.1342_8zotexr7.pdf

Gur, N., & Jackson, J. (2020). Procedure–content interaction in attitudes to law and in the value of the rule of law. In D. Meyerson, C. Mackenzie, & T. MacDermott (Eds.), *Procedural justice and relational theory: Empirical, philosophical, and legal perspectives* (pp. 111–140). Routledge. https://doi.org/10.4324/9780429317248-8

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, *108*(4), 814–834. https://doi.org/10.1037/0033-295X.108.4.814

Hannikainen, I. R., Tobia, K. P., de Almeida, G. F. C. F., Struchiner, N., Kneer, M., Bystranowski, P., Dranseika, V., Strohmaier, N., Bensinger, S., Dolinina, K., Janik, B., Lauraitytė, E., Laakasuo, M., Liefgreen, A., Neiders, I., Próchnicki, M., Rosas, A., Sundvall, J., & Żuradzki, T. (2022). Coordination and expertise foster legal textualism. *Proceedings of the National Academy of Sciences, USA*, *119*(44), Article e2206531119. https://doi.org/10.1073/pnas.2206531119

Hart, H. L. A. (1958). Positivism and the separation of law and morals. *Harvard Law Review*, *71*(4), 593–629. https://doi.org/10.2307/1338225

Hart, H. L. A. (1979). The new challenge to legal positivism. *Oxford Journal of Legal Studies*, *36*(3), 459–475. https://doi.org/10.1093/ojls/gqw021

Hart, H. M., Jr., & Sacks, A. M. (1994). *The legal process: Basic problems in the making and application of law*. Foundation Press.

Horowitz, I., & Kerr, N. (2001). Jury nullification: Legal and psychological perspectives. *Brooklyn Law Review*, *66*(4), 1207–1249.

Jaeger, B. (2017). *r2glmm* (R package Version 0.1.2) [Computer software]. GitHub. https://github.com/bcjaeger/r2glmm

Jones v. Minister for Justice and Equality, IECA 285 (Court of Appeal of Ireland 2019). https://www.courts.ie/acc/alfresco/31702c60-3d89-4ff8-a581-5ba6454fa1cf/2019_IECA_285_1.pdf/pdf

Kahan, D. (2010). Culture, cognition, and consent: Who perceives what, and why, in acquaintance-rape cases. *University of Pennsylvania Law Review*, *158*(3), 729–813. https://www.jstor.org/stable/20698345

Kennedy, D. (1976). Form and substance in private law adjudication. *Harvard Law Review*, *89*(8), 1685–1778. https://doi.org/10.2307/1340104

Kennett, J., & Fine, C. (2009). Will the real moral judgment please stand up? *Ethical Theory and Moral Practice*, *12*(1), 77–96. https://doi.org/10.1007/s10677-008-9136-4

Kirfel, L., & Hannikainen, I. R. (2023). Why blame the ostrich? Understanding culpability for willful ignorance. In K. Prochownik & S. Magen (Eds.), *Advances in experimental philosophy of law* (pp. 75–98). Bloomsbury Press.

Kirfel, L., & Phillips, J. (2023). The pervasive impact of ignorance. *Cognition*, *231*, Article 105316. https://doi.org/10.1016/j.cognition.2022.105316

Kneer, M., & Bourgeois-Gironde, S. (2017). Mens rea ascription, expertise and outcome effects: Professional judges surveyed. *Cognition*, *169*, 139–146. https://doi.org/10.1016/j.cognition.2017.08.008

Kneer, M., & Skoczeń, I. (2023). Outcome effects, moral luck and the hindsight bias. *Cognition*, *232*, Article 105258. https://doi.org/10.1016/j.cognition.2022.105258

Knobe, J. (2003). Intentional action and side effects in ordinary language. *Analysis*, *63*(3), 190–194. https://doi.org/10.1093/analys/63.3.190

Knobe, J., & Fraser, B. (2008). Causal judgments and moral judgment: Two experiments. In W. Sinnott-Armstrong (Ed.), *Moral psychology: Volume 2. The cognitive science of morality* (pp. 441–447). MIT Press.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects model. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

LaCosse, J., & Quintanilla, V. (2021). Empathy influences the interpretation of whether others have violated everyday indeterminate rules. *Law and Human Behavior*, *45*(4), 287–309. https://doi.org/10.1037/lhb0000456

Lewis, O., & Steinmo, S. (2012). How institutions evolve: Evolutionary theory and institutional change. *Polity*, *44*(3), 314–339. https://doi.org/10.1057/pol.2012.10

MacCormick, N., & Summers, R. (1991). *Interpreting statutes: A comparative study*. Routledge.

McBarnet, D., & Whelan, C. (1991). The elusive spirit of the law: Formalism and the struggle for legal control. *The Modern Law Review*, *54*(6), 848–873. https://doi.org/10.1111/j.1468-2230.1991.tb01854.x

McHugh, C., McGann, M., Igou, E. R., & Kinsella, E. L. (2017). Searching for moral dumbfounding: Identifying measurable indicators of moral dumbfounding. *Collabra. Psychology*, *3*(1), Article 23. https://doi.org/10.1525/collabra.79

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. https://www.jstor.org/stable/23216756

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), 133–142. https://doi.org/10.1111/j.2041-210x.2012.00261.x

Nance, D. (2006). Rules, standards, and the internal point of view. *Fordham Law Review*, *75*, 1287–1316.

Nobes, G., & Martin, J. W. (2022). They should have known better: The roles of negligence and outcome in moral judgements of accidental actions. *British Journal of Psychology*, *113*(2), 370–395. https://doi.org/10.1111/bjop.12536

Okonofua, J. A., & Eberhardt, J. L. (2015). Two strikes: Race and the disciplining of young students. *Psychological Science*, *26*(5), 617–624. https://doi.org/10.1177/0956797615570365

Patil, I., & Trémolière, B. (2021). Reasoning supports forgiving accidental harms. *Scientific Reports*, *11*(1), Article 14418. https://doi.org/10.1038/s41598-021-93908-z

Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, *70*, 153–163. https://doi.org/10.1016/j.jesp.2017.01.006

Peter-Hagene, L., & Bottoms, B. (2017). Attitudes, anger, and nullification instructions influence jurors' verdicts in euthanasia cases. *Psychology, Crime & Law*, *23*(10), 983–1009. https://doi.org/10.1080/1068316X.2017.1351967

Peter-Hagene, L. C., & Ratliff, C. L. (2021). When jurors' moral judgments result in jury nullification: Moral outrage at the law as a mediator of euthanasia attitudes on verdicts. *Psychiatry, Psychology and Law*, *28*(1), 27–49. https://doi.org/10.1080/13218719.2020.1751741

Phillips, J., & Cushman, F. (2017). Morality constrains the default representation of what is possible. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(18), 4649–4654. https://doi.org/10.1073/pnas.1619717114

Pound, R. (1959). *Jurisprudence*. West Publishing.

Raz, J. (1999). *Practical reason and norms*. Oxford University Press. https://doi.org/10.1093/acprof:Oso/9780198268345.001.0001

Richard, F. D., Bond, C. F., Jr., & Stokes-Zoota, J. J. (2003). One hundred years of social psychology quantitatively described. *Review of General Psychology*, *7*(4), 331–363. https://doi.org/10.1037/1089-2680.7.4.331

Richardson, L. S., & Goff, P. (2012). Self-defense and the suspicion heuristic. *Iowa Law Review*, *98*, 293–336.

Roberts, J. (2005). *Confirmation hearing on the nomination of John G. Roberts, Jr. to be Chief Justice of the United States: Hearings before the Committee on the Judiciary, United States Senate, One Hundred Ninth Congress, First Session, September 12–15, 2005*. U.S. Government Printing Office.

Salerno, J. M., & Peter-Hagene, L. C. (2013). The interactive effect of anger and disgust on moral outrage and judgments. *Psychological Science*, *24*(10), 2069–2078. https://doi.org/10.1177/0956797613486988

Schauer, F. (2005). The social construction of the concept of law: A reply to Julie Dickson. *Oxford Journal of Legal Studies*, *25*(3), 493–501. https://doi.org/10.1093/ojls/gqi024

Schauer, F. (2009). Institutions and the concept of law: A reply to Ronald Dworkin (with some help from Neil MacCormick). In M. del Mar & Z. Bankowski (Eds.), *Law as institutional normative order* (pp. 35–45). Routledge.

Scheflin, A. (1972). Jury nullification: The right to say no. *Southern California Law Review*, *45*, 168–226.

Schlag, P. (1985). Rules and standards. *UCLA Law Review*, *33*, 379–430.

Schwartz, F., Djeriouat, H., & Trémolière, B. (2022). Agents' moral character shapes people's moral evaluations of accidental harm transgressions. *Journal of Experimental Social Psychology*, *102*, Article 104378. https://doi.org/10.1016/j.jesp.2022.104378

Shapiro, S. J. (2011). *Legality*. Harvard University Press. https://doi.org/10.2307/j.ctvjnrsd5

Shenhav, A., & Greene, J. D. (2010). Moral judgments recruit domain-general valuation mechanisms to integrate representations of probability and magnitude. *Neuron*, *67*(4), 667–677. https://doi.org/10.1016/j.neuron.2010.07.020

Sinclair, M. B. W. (1997). Legislative intent: Fact or fabrication. *New York Law School Law Review. New York Law School*, *41*, 1329–1389.

Sinnott-Armstrong, W. (2005). Word meaning in legal interpretation. *The San Diego Law Review*, *42*, 465–492.

Skorinko, J. L., Laurent, S., Bountress, K., Nyein, K. P., & Kuckuck, D. (2014). Effects of perspective taking on courtroom decisions. *Journal of Applied Social Psychology*, *44*(4), 303–318. https://doi.org/10.1111/jasp.12222

Spooner, L. (1847). *The unconstitutionality of slavery*. Bela Marsh.

Sterling, J., & Moore, W. (1987). Weber's analysis of legal rationalization: A critique and constructive modification. *Sociological Forum*, *2*(1), 67–89. https://doi.org/10.1007/BF01107894

Struchiner, N., Hannikainen, I. R., & de Almeida, G. F. C. F. (2020). An experimental guide to vehicles in the park. *Judgment and Decision Making*, *15*(3), 312–329. https://doi.org/10.1017/S1930297500007130

Sullivan, K. (1992). The justices of rules and standards. *Harvard Law Review*, *106*, 22–123. https://doi.org/10.2307/1341533

Suter, R. S., & Hertwig, R. (2011). Time and moral judgment. *Cognition*, *119*(3), 454–458. https://doi.org/10.1016/j.cognition.2011.01.018

Teoh, Y. Y., Yao, Z., Cunningham, W. A., & Hutcherson, C. A. (2020). Attentional priorities drive effects of time pressure on altruistic choice. *Nature Communications*, *11*(1), Article 3534. https://doi.org/10.1038/s41467-020-17326-x

Tinghög, G., Andersson, D., Bonn, C., Johannesson, M., Kirchler, M., Koppel, L., & Västfjäll, D. (2016). Intuition and moral decision-making—The effect of time pressure and cognitive load on moral judgment and altruistic behavior. *PLOS ONE*, *11*(10), Article e0164012. https://doi.org/10.1371/journal.pone.0164012

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). mediation: R package for causal mediation analysis. *Journal of Statistical Software*, *59*(5), 1–38. https://doi.org/10.18637/jss.v059.i05

Tobia, K. P. (2018). How people judge what is reasonable. *Alabama Law Review*, *70*(2), 293–359.

Turri, J. (2019). Excuse validation: A cross-cultural study. *Cognitive Science*, *43*(8), Article e12748. https://doi.org/10.1111/cogs.12748

Turri, J., & Blouw, P. (2015). Excuse validation: A study in rule-breaking. *Philosophical Studies*, *172*(3), 615–634. https://doi.org/10.1007/s11098-014-0322-z

United States v. Carroll Towing Co., 159 F.2d 169 (2d. Cir. 1947). https://law.justia.com/cases/federal/appellate-courts/F2/159/169/1565896/

Weitzer, R., & Brunson, R. (2015). Policing different racial groups in the United States. *Cahiers Politiestudies*, *35*, 129–145.

Wylie, J., & Gantman, A. (2023). Doesn't everybody jaywalk? On codified rules that are seldom followed and selectively punished. *Cognition*, *231*, Article 105323. https://doi.org/10.1016/j.cognition.2022.105323

Young, L., Nichols, S., & Saxe, R. (2010). Investigating the neural and cognitive basis of moral luck: It's not what you do but what you know. *Review of Philosophy and Psychology*, *1*(3), 333–349. https://doi.org/10.1007/s13164-010-0027-y

Received March 31, 2022
Revision received January 27, 2023
Accepted January 29, 2023 ■