Review of Philosophy and Psychology
https://doi.org/10.1007/s13164-023-00674-0

Justice before Expediency: Robust Intuitive Concern for Rights Protection in Criminalization Decisions

Piotr Bystranowski[1] · **Ivar Rodríguez Hannikainen**[2]

Accepted: 10 January 2023

Abstract

The notion that a false positive (false conviction) is worse than a false negative (false acquittal) is a deep-seated commitment in the theory of criminal law. Its most illustrious formulation, the so-called Blackstone's ratio, affirms that "it is better that ten guilty persons escape than that one innocent suffer". Are people's evaluations of criminal statutes consitent with this tenet of the Western legal tradition? To answer this question, we conducted three experiments (total $N = 2492$) investigating how people reason about a particular class of offenses—proxy crimes—known to vary in their specificity and sensitivity in predicting actual crime. By manipulating the extent to which proxy crimes convict the innocent and acquit those guilty of a target offense, we uncovered evidence that attitudes toward proxy criminalization depend primarily on its propensity toward false positives, with false negatives exerting a substantially weaker effect. This tendency arose across multiple experimental conditions—whether we matched the rates of false positives and false negatives or their frequencies, whether information was presented visually or numerically, and whether decisions were made under time pressure or after a forced delay—and was unrelated to participants' probability literacy or their professed views on the purpose of criminal punishment. Despite the observed inattentiveness to false negatives, when asked to justify their decisions, participants retrospectively supported their judgments by highlighting the proxy crime's efficacy (or inefficacy) in combating crime. These results reveal a striking inconsistency: people favor criminal policies that protect the rights of the innocent, but report comparable concern for their expediency in fighting crime.

✉ Piotr Bystranowski
piotr.bystranowski@uj.edu.pl

[1] Interdisciplinary Centre for Ethics & Institute of Philosophy, Jagiellonian University, Grodzka 52, 31-044 Kraków, Poland

[2] Department of Philosophy I, University of Granada, Granada, Spain

Published online: 28 January 2023

1 Introduction

It was a felony in eighteenth century England for a mother to conceal the birth of her child, as such behavior indicated that she intended to kill, or had already killed, the child (Bentham 1887). Similarly, many countries now punish citizens for merely traveling to areas under the control of terrorist groups, on the presumption that these individuals are involved in acts of terrorism (De Guttry et al. 2016).

Why criminalize these *proxy* conducts that are only probabilistically linked to criminal acts, and not wrongful in themselves? One reason is that actual criminal misconduct can be hard for law enforcement to observe—since offenders are motivated to avoid punishment by concealing their wrongdoing. Second, even when criminal conduct could in principle be observed, it may be prohibitively costly (or even impossible) to prove it beyond a reasonable doubt in order to secure criminal conviction. *Proxy crimes* offer a practical solution to this problem.[1]

On a theoretical plane, proxy crimes help to reveal the principles that guide people's decisions about criminal policy. Because proxy crimes target primary wrongdoing *indirectly*, they are both overinclusive (there are instances of the proxy conduct that do not incur in the primary wrongdoing) and underinclusive (there are instances of the primary wrongdoing that are not associated with the proxy conduct). The scope of that over- and under-inclusion is arguably the main factor that should determine the normative assessment of a given proxy offense. On one hand, proxy crimes can be evaluated on the basis of their effectiveness in curtailing crime: that is, proxy offenses are acceptable insofar as they maximize *sensitivity* (i.e., they help convict many actual criminals) without incurring too many costs on the innocent.

On the other hand, proxy crimes can be evaluated on the basis of their justice, or respect for the innocent's rights: that is, proxy crimes are acceptable insofar as they maximize *specificity* and refrain from convicting citizens who are innocent of the primary wrongdoing. These considerations are essential for scholars who think that punishment is only justified by the defendant's criminal *actus reus* (Duff et al. 2007; Picinali 2018)—an act which cannot be replaced by mere correlates of crime (Alexander and Ferzan 2009; Duff 1997). Doing so interferes with fundamental principles of criminal justice, such as the presumption of innocence (Tomlin 2013).

The way proxy crimes are drafted is directly linked to a central value judgment that any criminal law system has to make: To what extent the expecidency in curtailing crime should be constrained by the respect towards the rights of the innocent. While legal scholarship has traditionally addressed this trade-off by some reference to the so-called Blackstone's ratio (it is better "that ten guilty persons escape than that one innocent suffer"; Blackstone 1765), little is known about the way regular people approach such dilemmas.

[1] While proxy crimes might appear at first to be a peculiar and marginal phenomenon, many criminal law scholars have noticed that they constitute a vast and consistently growing category of criminal offenses in modern legal systems (McAdams 2005; Stuntz 2001) and one of the leading causes of the phenomenon of overcriminalization (Husak 2008; Lee 2021). They are related to, although distinct from, the class of *phantom rules* (Wylie and Gantman 2023).

Related research on punishment (Carlsmith 2008; Carlsmith et al. 2002; Darley et al. 2000; McFatter 1982; Sharp and Otto 1910) has already demonstrated that non-consequentialist—that is, primarily retributivist—motives guide people's sentencing recommendations. Furthermore, despite evincing retributivist motives, people decry retributivist rationales (aimed at punishing the guilty for their past wrongs) and gravitate toward consequentialist justifications for punishment (aimed at deterring future crime) instead (Carlsmith 2008). This discrepancy between participants' tacit judgments and their professed convictions has been documented pervasively in previous studies on moral cognition (Haidt 2001).

In the present work, we examine the principles that guide people's decisions about *criminalization*—that is, about whether various criminal statutes should be adopted in the first place. To pursue this question, we focus on proxy crimes (Bystranowski and Mungan 2022; McAdams 2005), and evaluate the extent to which factors crucial to expediency and justice shape people's support for proxy criminal policies. For instance, to what extent does people's support for the aforementioned travel ban depend on whether it helps to apprehend terrorists *and* on whether it preserves innocent travelers' right to transit?

In our studies, we vary the rates of false positives and false negatives and investigate whether participants' attitudes toward proxy criminal statutes are influenced by these manipulations. If participants display efficacy-centered attitudes toward criminalization, their evaluation of proxy crimes ought to depend on both the false positive and false negative rates. If instead people demonstrate rights-protection principles, their approval of proxy crimes ought to depend primarily on the rate of false positives.

Whether such intuitive principles are introspectively recognized and expressly endorsed is a separate question. Growing evidence under the guise of social intuitionism (Haidt 2001) highlights how the eliciting factors that shape people's intuitive judgments bear little relation to the justifying reasons people subsequently provide (Carlsmith et al. 2002; Almagro et al. 2022). In the present research, we explore the possibility that attitudes toward criminalization exhibit a similar discrepancy between the factors that shape intuitive preferences regarding criminalization, and the reasoning that people provide post hoc.

Data, analysis scripts and materials are openly accessible on the *Open Science Framework* at: https://osf.io/tn6j8/. Statistical analyses were conducted in *R* version 4.1.2 using the *lme4* and *emmeans* packages. Our primary analysis are mixed-effects models with random intercepts of participant and scenario, for which we report *F* tests using the Kenward-Roger approximation to the degrees of freedom.

2 Study 1

In Study 1, we investigate the factors that shape people's attitudes toward proxy criminal statutes. Participants evaluated a series of proposed proxy crimes devised to address a target wrongdoing (see Table 1). We manipulated the rates of false positives and false negatives (as well as the base rate of crime) for each proxy criminal statute, and evaluated their impact on people's evaluations. In addition, we examined people's justifications for their evaluations of each proxy crime through multi-item measures.

2.1 Methods

2.1.1 Participants

Data were collected through the Prolific Academic crowdsourcing platform in two waves (May 2020 and March 2021). For simplicity, we report the results of both waves in aggregate. Participants were (i) native English speakers with (ii) at least a high school/secondary education. For each iteration of the study, we recruited 400 participants (498 women, age: $M = 34.2$ years, $SD = 12.9$). We did not apply any exclusion criteria or exclude any observations.

The target sample size was decided before any data analysis on the basis of funding availability. A total sample size of 800 participants afforded excellent statistical power (95%) to detect a small-to-medium correlation ($r = .18$; smaller than the average effect in social psychology, see Richard et al. 2003) when α equals .05. We report all measures and manipulations.

2.1.2 Materials

Participants viewed four short vignettes in a randomized order. The introduction to each vignette described an unlawful activity occurring in a hypothetical community:

> *A dangerous cult is gaining popularity in the country of Zubrovka. Most of its followers have settled outside state control, in the Western Desert, where they fight the government army and from where they send assassins that spread fear across the country. Although involvement in the illegal activities of the cult is a serious crime, growing numbers of young people are joining the cult, as it is almost impossible to prove that they are involved in criminal wrongdoing in the Western Desert.*

Next, participants learned that lawmakers were considering whether to pass a proxy criminal statute:

> *The Parliament of Zubrovka considers passing a new offense to make prosecution more effective: The Anti-Cult Act. The Anti-Cult Act would make it a crime to enter Zubrovka from the Western Desert. These days, most people leaving the Western Desert have been involved in the illegal activities of the cult. The Western Desert used to be a popular getaway for residents of Zubrovka, but travel in the region has been steadily declining due to the risk posed by the cult. So, anyone traveling to Zubrovka from the Western Desert is likely a member of the cult.*

The proxy conduct either causally preceded the target offense (i.e., traveling to the Western Desert) or resulted from the target offense (i.e., traveling from the Western Desert). Table 1 lists each target offense and both corresponding proxy behaviors.

Table 1 Stimuli: primary offenses and corresponding proxy conducts

Target offense	Precursor	Consequent
Poaching	Possession of swift bullets	Selling bear skin
Doping	Possession of a syringe	Taking a drug that masks the presence of the illicit substance
Fraud	Possession of a marked deck of cards	Possession of over $500 in cash while being a vagrant
Terrorism	Possession of a bomb-construction manual	Wearing head-covering in public after an explosion
Cult *	Travelling *to* a terrorist-controlled territory	Travelling *from* a terrorist-controlled territory

*: The Cult scenario was only employed in Study 2

Participants then learned of the proxy criminal policy's *specificity* and *sensitivity* in predicting the target wrongdoing, as well as the *base rate* of the target wrongdoing in the community. These values were displayed visually in the form of a dot plot (see Fig. 1), purportedly part of a report commissioned by the lawmakers to learn more about the unlawful activity. The *visual display* was presented on a separate screen, and participants could not advance to the following page until 8 seconds had elapsed. In the visual display, criminals and innocents were lateralized and engaging in versus refraining from the proxy conduct was color-coded (see Fig. 1).

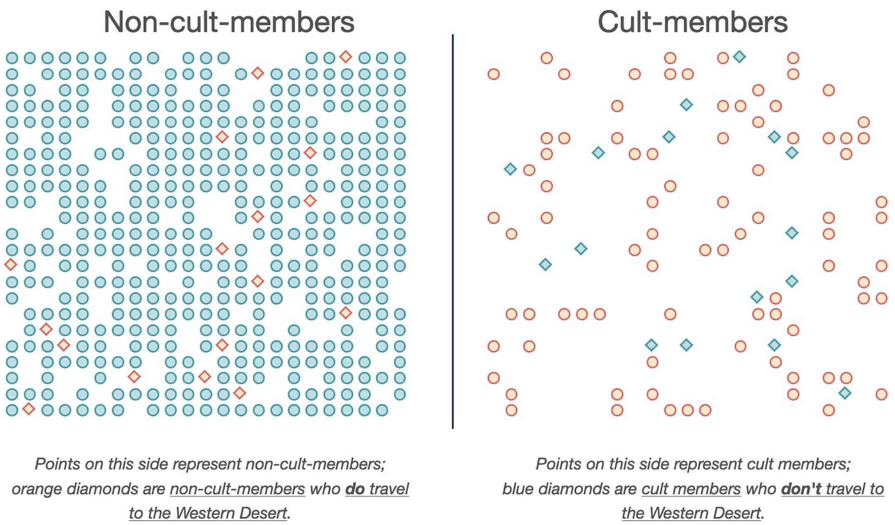
To generate the four frequencies (true and false negatives, and true and false positives), we defined three parameters:

1) the *base rate*: the proportion of the entire population that engages in the primary wrongdoing,
2) the *false negative rate*: the proportion of criminals who do not engage in the proxy conduct, and
3) the *false positive rate*: the proportion of innocents who do engage in the proxy conduct.

These parameters were randomly sampled on each trial from uniform distributions. The base rate was allowed to range between 10% and 30% of the population. The false negative and false positive rates were allowed to range between 1% and 19% (of the criminals and innocents, respectively). Thus, the frequency of false negatives would equal *population size* (which was held constant at 526) × *base rate* × *false negative rate*, rounded to the nearest integer.

2.1.3 Procedure

After providing informed consent, participants completed the *Training Task*, the *Experimental Block*, and the *Individual Difference Measures* in a fixed order.



Non-cult-members

Cult-members

Points on this side represent non-cult-members; orange diamonds are non-cult-members who ***do*** *travel to the Western Desert.*

Points on this side represent cult members; blue diamonds are cult members who ***don't*** *travel to the Western Desert.*

Fig. 1 Sample display

Training Task Participants first saw two example displays showing the prevalence of an undesirable activity (e.g., alcohol abuse) and its associated "symptom" in a community (e.g., having bloodshot eyes). They were then asked five true or false questions, each of which required an inference from the example displays (e.g., 'A person with bloodshot eyes is more likely than not to abuse alcohol'). The purpose of this task was for participants to familiarize with the visual display and practice their interpretation. We calculated the sum score of correct responses on the training task ($M = 3.36$, $SD = 1.17$) as an individual difference measure of *probability literacy*.

Experimental Block In a balanced incomplete block design, each participant was assigned to a block in which they viewed two precursor and two consequent proxy criminal policies. In each block, participants viewed either the Precursor or Consequent variant of each of four offenses and judged the battery of four cases in a randomized order. In Online Supplement 1, we report the effect of the precursor versus consequent manipulation on participants' responses.

Each case was composed of a sequence of slides: the *vignette* (describing the primary offense and the proposed proxy crime), the visual *display*, followed by the *decision*, and *justification* slides. On the *decision* slide, participants were asked whether the proxy criminal policy should be passed. Then, on the *justification* slide, participants evaluated the proxy crime's effectiveness and justice (see Measures subsection).

Individual Difference Measures After the battery of four vignettes, participants completed two individual difference measures: the 6-item *Punitiveness Questionnaire* (Kemme et al. 2014), and an adaptation of the retribution/deterrence motives task introduced by Carlsmith and colleagues (Carlsmith et al. 2002). Finally, participants optionally provided the following demographic information: their gender, age, educational attainment, religiosity, and political orientation.

2.1.4 Measures

For each scenario and participant, we recorded the following three independent measures: *t*he *base rate*, the *false positive* rate*,* and the *false negative* rate. The dependent measure on each trial was *approval* (i.e., whether the proxy crime should be adopted on a 7-point Likert scale anchored at 1: ‘Certainly not’ and 7: ‘Certainly’).

Participants also completed a six-item assessment of each proxy crime’s effectiveness (whether the proxy crime helps apprehend offenders/prevents wrongdoing/deters potential offenders) and justice (whether it punishes/affects the lives of/is unfair toward innocent citizens). The items were rated on 7-point Likert scales, anchored at 1: ‘Strongly Disagree’ and 7: ‘Strongly Agree’. We submitted these responses to a maximum-likelihood factor analysis with varimax rotation. Two factors exhibited factor loadings above one (2.18, and 2.01, with the 3rd factor dropping to 0.18), and explained 70% of the variance in evaluations. Every item loaded onto a single factor, as hypothesized, with factor loadings > .73 (and failed to load onto the other factor, with the absolute value of every factor loading < .24). Thus, we calculated two additional measures per trial: (1) *effectiveness,* the three-item average (Cronbach’s $\alpha = .88$) of whether the proxy crime effectively combats crime; and (2) *justice,* the three-item average (Cronbach’s $\alpha = .86$) of whether the proxy crime safeguards the rights of innocent citizens.

2.2 Results

Summary statistics for approval, justice, and effectiveness are reported in Table 2. In the aggregate, participants tended to approve of the proxy crimes ($M = 4.92$, 95% CI [4.86, 4.97]), and viewed them as effective ($M = 5.05$, 95% CI [5.01, 5.10]), yet unjust ($M = 3.85$, 95% CI [3.80, 3.90]).

In the analyses below, we examine whether participants’ approval judgments were shaped by rates of false positives and/or false negatives (while accounting for variation in the base rate of crime). To this end, we enter approval as the dependent measure in a mixed-effects model, with random effects of participant and scenario, and rates of false positives, false negatives, and the base rate as fixed effects. We then repeat this procedure with the secondary dependent measures of perceived effectiveness and justice.

2.2.1 Effects of False Positive and False Negative Rates

Approval of a proxy crime depended on the rate of false positives ($B = -4.66$, 95% CI [−5.63, −3.69], $t = -9.39$, $p < .001$), but not on the rate of false negatives ($B = 0.21$, 95% CI [−0.75, 1.18], $t = 0.44$, $p = .66$), or on the base rate of crime ($B = 0.53$, 95% CI [−0.36, 1.42], $t = 1.17$, $p = .24$; see Fig. 2).

Table 2 Descriptive statistics: Study 1

Scenario	Condition	N	Approval *M* (*SD*)	Justice *M* (*SD*)	Effectiveness *M* (*SD*)
Poaching	Precursor	400	5.42 (1.55)	4.28 (1.32)	5.32 (1.15)
	Consequent	400	5.56 (1.57)	4.29 (1.39)	5.46 (1.19)
Doping	Precursor	393	4.46 (1.70)	3.34 (1.46)	5.07 (1.24)
	Consequent	407	4.85 (1.58)	3.37 (1.51)	5.30 (1.21)
Fraud	Precursor	400	4.90 (1.75)	3.93 (1.45)	5.07 (1.29)
	Consequent	400	4.15 (1.71)	3.35 (1.32)	4.65 (1.24)
Terrorism	Precursor	407	5.42 (1.48)	4.21 (1.49)	5.09 (1.32)
	Consequent	393	4.54 (1.78)	3.64 (1.50)	4.44 (1.49)

2.2.2 Effectiveness and Justice Evaluations

A corresponding model of justice evaluations indicated that false positives negatively predicted justice, $B = -7.11$, 95% CI [−7.91, −6.30], $t = -17.26$, $p < .001$. Meanwhile, we observed no effect of the rate of false negativess on perceptions of justice ($B = 0.39$, 95% CI [−0.41, 1.19], $t = 0.96$, $p = .34$) and inconclusive evidence of a weak effect of variation in the base rate ($B = 0.74$, 95% CI [−0.00, 1.48], $t = 1.95$, $p = .051$). Thus, as expected, the more a proxy crime falsely convicted innocent citizens, the more likely participants were to view it as unjust.

Next, we examined the determinants of proxy crimes' perceived effectiveness. We observed no effect of the base rate of crime on effectiveness, $B = 0.28$, 95% CI [−0.39, 0.94], $t = 0.82$, $p = .41$, and some inconclusive evidence of a weak effect of the rate of false negatives, $B = -0.67$, 95% CI [−1.38, 0.05], $t = -1.83$, $p = .067$. In contrast, reported effectiveness depended on the rate of false positives, $B = -2.81$, 95% CI [−3.53, −2.09], $t = -7.63$, $p < .001$. In other words, proxy crimes were perceived as effective in apprehending criminals not when few criminals were acquitted, but rather when few innocent citizens were convicted.

We therefore reasoned that reports of effectiveness served to justify participants' decisions (i.e., to approve or reject the proxy criminal statute). If so, the influence of false positives on effectiveness may reflect an indirect effect via approval. Indeed, approval and effectiveness were strongly related, $B = 0.44$, 95% CI [0.43, 0.47], $t = 39.89$, $p < .001$. Furthermore, entering approval as an additional predictor in the model of effectiveness substantially weakened the influence of false positives ($B = -0.76$, 95% CI [−1.36, −0.16], $t = -2.48$, $p = .013$).

2.3 Discussion

Study 1 provided evidence that false positives shape attitudes toward criminal policy—whereas false negatives have a limited effect, if any. Participants were greatly concerned with the risk of convicting innocent citizens, but largely inattentive to the risk of acquitting criminals.

Despite this fact, when asked to provide reasons in support of their prior decision, supporters claimed that the policies were effective in combating crime while detractors

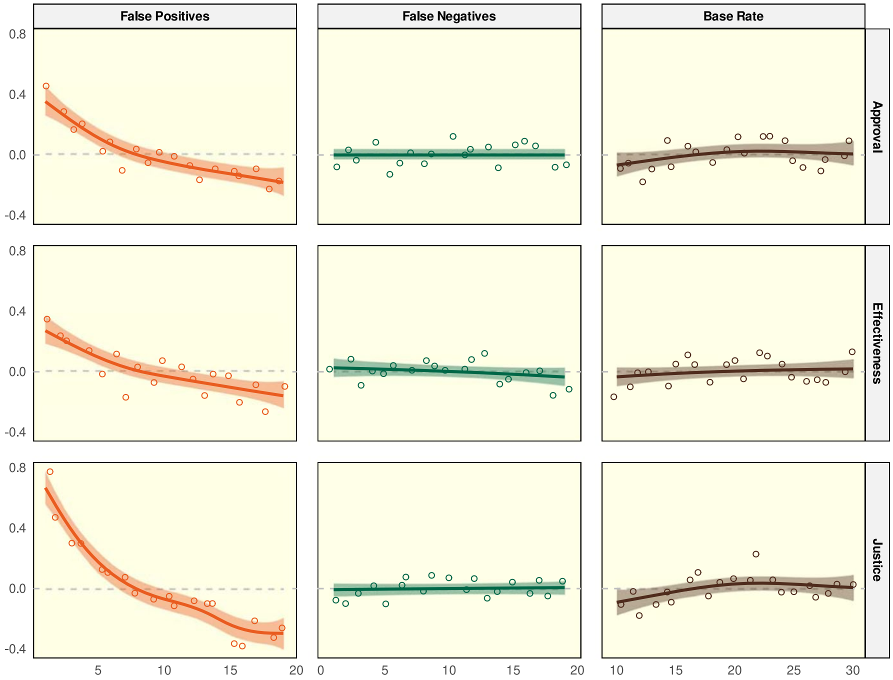


Fig. 2 Study 1: Standardized (i.e., z-scored) evaluations of the proxy criminal policies by false positive and acquittal rates, and by the base rate of crime. Each plot displays (locally estimated) curve fit and mean values in percent increments

claimed that they were ineffective. As such, the perceived effectiveness of a proxy crime was weakly related to the proportion of criminals which it helped to apprehend, and primarily depended on the rate at which innocent citizens were falsely convicted. In this regard, Study 1 points toward a dissociation between the factors that shape participants' judgments and their subsequent justifications: Considerations of justice strongly impacted decisions regarding proxy criminal statutes, whereas considerations of expediency were brought to bear to justify them.

3 Study 2

Study 1 documented a dominant role of false positives in shaping criminalization preferences. In Study 2, we examine the cognitive basis of this effect through multiple experimental manipulations. First, we impose time pressure in order to understand whether participants manifest an emphasis on false positives even when faced with limited time to decide. Such a result would suggest that the emphasis on false positives reflects a relatively intuitive and automatic cognitive process. Second, we investigate whether encouraging participants to reflect—by (i) providing additional time to decide, and/or (ii) previously assessing the proxy crime's effectiveness and justice—weakens the effect of false positives and/or magnifies the effect of false negatives on approval. Third, we also compare participants' criminalization decisions when (iii) information is presented in visual versus numerical form–on the assumption that numerical summary data would facilitate participants' calculations of the probative value of each proxy crime.

Finally, in Study 2 we addressed certain limitations in our design of Study 1: Specifically, the visual display in Study 1 used a fixed lateralization and color scheme, which could have influenced participants' allocation of attention. In Study 2, we randomized the lateralization and regimented the color scheme to rule out these potential effects.

3.1 Methods

3.1.1 Participants

751 participants (373 women, age: $M=36.2$ years, $SD=12.4$) were recruited through the Prolific Academic crowdsourcing platform in November 2021. Participants who failed a simple attention check were not allowed to complete the survey. Participants were (i) native English speakers with (ii) at least a high school/secondary education. The target sample size was decided before any data analysis based on a rule of thumb: We sought to recruit at least 120 participants per condition in a 3×2 between-subjects design (i.e., 720).

3.1.2 Materials

Participants viewed five short vignettes (see Table 1) in a randomized order. Each participant was presented with two or three vignettes in the Precursor version, with the remaining vignettes in the Consequent version.

As in Study 1, participants received information about the probative value of each proxy conduct. In Study 2, we manipulated whether the information was displayed in visual form (as in Study 1; with a uniform color scheme while randomizing lateralization across participants) or numerical form. Participants in the numerical conditions received the information in the form of a confusion matrix. For example, in the Cult scenario (where the target wrongdoing is participation in a dangerous cult, and the proxy offense is traveling *to* or *from* the Western Desert), participants might receive the following information:

> *There are 73 cult members and 448 non-members.*
> *69 cult members travel to the Western Desert.*
> *4 cult members don't travel to the Western Desert.*
> *394 non-members don't travel to the Western Desert.*
> *54 non-members travel to the Western Desert.*

3.1.3 Procedure

In a 2 (mode: Visual, Numerical) × 3 (condition: Time Pressure, Forced Delay, Deliberation) between-subjects design, participants were randomly assigned to one of six conditions. As in Study 1, participants in every condition completed a *Training Task*, the *Experimental Block*, and the *Individual Difference Measures* in a fixed order.

Training Task The training task was composed of either visual displays or numerical information, depending on condition assignment. In the Visual mode, the training task was just as in Study 1. In the Numerical mode, participants were presented with corresponding confusion matrices. We again defined *probability literacy* as the sum score of correct answers to five true or false questions. Probability literacy was slightly higher in the numerical mode ($M = 3.54$, $SD = 1.07$) than the visual mode ($M = 3.39$, $SD = 1.09$), Welch's $t(748) = 1.94$, $p = .052$, Cohen's $d = 0.14$).

Experimental Block Next, participants viewed five scenarios in a randomized order. Participants in the Forced Delay condition followed the procedure as in Study 1, with the vignette, display (visual or numerical according to condition), decision and justification slides in sequence. The Time Pressure condition differed from the Forced Delay condition in that the display slide was presented for 12 seconds and then automatically advanced. In the Deliberation condition, the decision and justification questions were presented on the same screen as the display, with the justification items before the decision question.

Individual Difference Measures As in Study 1, participants completed the individual difference measures and provided demographic information after the experimental portion of the study.

3.1.4 Measures

As in Study 1, for each scenario and participant, we recorded the three independent measures (*crime rate*, *false positives, false negatives*), and three dependent measures: *approval*, *effectiveness* (Cronbach's $\alpha = .89$), and *justice* (Cronbach's $\alpha = .87$).

3.2 Results

Below we report a series of mixed-effects models with scenarios and participants as crossed random effects. In the fixed effects portion of the model, we enter *mode* of presentation, and two dummy-codes reflecting whether the deliberation *nudge* was present or absent, whether *time pressure* was present or absent, and the two-way interactions between both dummy variables and presentation mode.

3.2.1 Effects of Condition and Presentation Mode

We observed no main effects of mode of presentation, deliberation nudge, time pressure, or any two-way interaction effects in our primary model of approval, all Fs < 1.41, ps $> .24$. Pairwise comparisons revealed no significant differences between conditions in the Numerical mode, ps $> .47$, or the Visual mode, ps $> .72$.

The intercept-only model revealed that the grand mean ($M = 4.97$) significantly differed from the midpoint, $t = 6.74$, $p < .001$, indicating that participants tended to approve of the proxy criminal policies across conditions.

A corresponding model of justice evaluations revealed a main effect of mode of presentation, $F_{(1,\ 745)} = 6.24$, $p = .013$, and no other effects, Fs < 0.66, ps $> .42$. This main effect reflected higher justice evaluations in the Visual display mode ($M = 4.00$, 95% CI [3.63, 4.38]) than the Numerical summary mode ($M = 3.83$, 95% CI [3.46, 4.21]), $t = 2.46$, $p = .014$.

In a model of effectiveness judgments, no significant effects were observed, all Fs < 0.04, ps $> .84$. The intercept-only model revealed that the grand mean ($M = 5.15$) significantly differed from the midpoint, $t = 11.95$, $p < .001$, indicating that participants tended to view the proxy criminal policies as effective across conditions.

3.2.2 Effects of False Positive and False Negative Rates

Next, we examined whether our continuous manipulations of the false positive and false negative rates, as well as of the base rate of crime, influenced participants' evaluations. As in Study 1, we observed a strong main effect of false positives on approval, $F_{(1,3435)} = 158.2$, $p < .001$. False positives reduced approval of proxy criminal policies across conditions, $B = -5.48$, 95% CI [−6.33, −4.63], $t = -12.66$, $p < .001$–and the effect was robust across presentation modes and conditions, $-6.39 < B\text{s} < -4.81$, $-6.12 < t\text{s} < -4.45$, $p\text{s} < .001$.

This time, however, we also observed significant (albeit weak) effects of false negatives, $F_{(1,\ 3397)} = 6.20$, $p = .012$, and the base rate, $F_{(1,\ 3411)} = 4.91$, $p = .027$. False negatives reduced approval of proxy criminal policies ($B = -1.07$, 95% CI [−1.92, −0.22], $t = -2.46$, $p = .014$) while increases in the base rate of crime promoted approval, $B = 0.87$, 95% CI [0.11, 1.62], $t = 2.24$, $p = .025$ (see Fig. 3).

Furthermore, the influence of false negatives varied across conditions–which was reflected by a non-significant three-way interaction with time pressure and mode, $F_{(1,\ 3408)} = 3.81$, $p = .051$. In the Numerical mode, time to reflect did not moderate the impact of false negatives on approval, $B = 1.26$, $t = 0.83$, $p = .41$. Furthermore, the simple effect of false negatives was non-significant in both the delayed ($B = -0.44$, 95% CI [−1.89, 1.00]) and speeded ($B = -1.71$, 95% CI [−4.34, 0.92]) conditions. However, in the Visual mode, the forced delay *did* strengthen the effect of false negatives on approval, $B = -2.86$, $t = -1.96$, $p = .050$. Specifically, false negatives reduced approval in the delayed condition ($B = -2.26$, 95% CI [−3.76, −0.76]) but had no effect under time pressure ($B = 0.60$, 95% CI [−1.92, 3.11]).

3.2.3 Effectiveness and Justice Evaluations

A model of justice evaluations revealed effects of false positives, $F_{(1,\ 3397)} = 327.9$, $p < .001$, but also false negatives, $F_{(1,\ 3360)} = 7.50$, $p = .006$, and the base rate, $F_{(1,\ 3374)} = 7.73$, $p = .005$. False positives negatively predicted justice evaluations, $B = -6.84$, 95% CI [−7.57, −6.10], $t = -18.21$, $p < .001$. This time, unlike Study 1, false

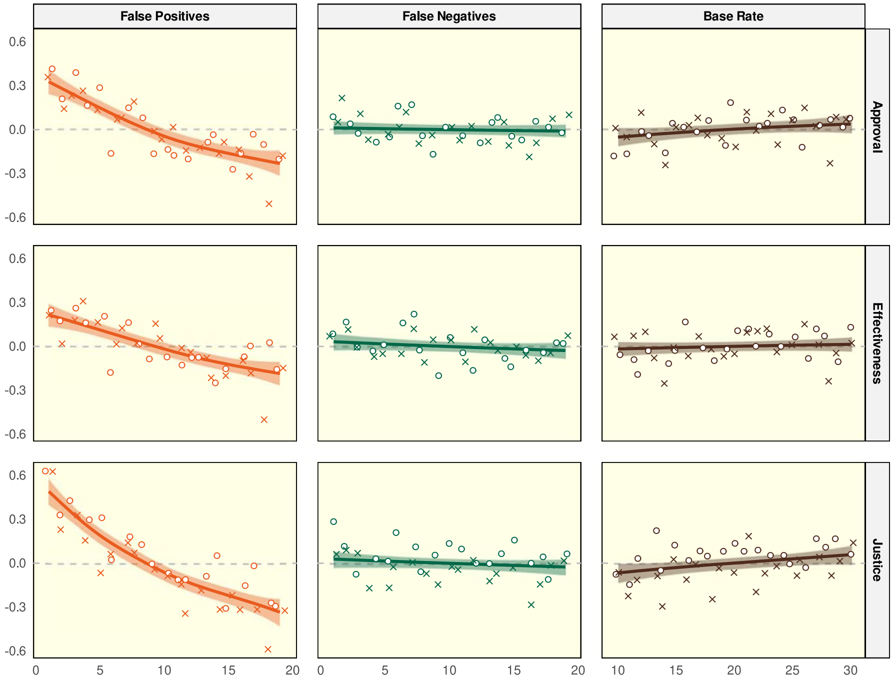


Fig. 3 Study 2: Standardized (i.e., z-scored) evaluations of the proxy criminal policies by false positive and false negative rates, and by the base rate of crime. Each plot displays (locally estimated) curve fit and mean values for each mode of presentation (Numerical: crosses; Visual: circles) in percent increments

negatives also rendered proxy crimes more unjust, $B = -1.04$, 95% CI [−1.77, −0.30], $t = -2.76$, $p = .006$. Additionally, increases in the base rate of crime rendered proxy crimes more just, $B = 0.92$, 95% CI [0.26, 1.58], $t = 2.75$, $p = .006$. Importantly, false negatives and crime rates had a weaker influence on perceptions of justice than did false positives.

A model of effectiveness evaluations revealed effects of both false positives, $F_{(1,\ 3319)} = 82.05$, $p < .001$, and false negatives, $F_{(1,\ 3288)} = 10.38$, $p = .001$–but not of the base rate, $F_{(1,\ 3299)} = 2.09$, $p = .15$. As in Study 1, false positives negatively predicted effectiveness, $B = -2.96$, 95% CI [−3.60, −2.33], $t = -9.13$, $p < .001$.

The main effect of false negatives on effectiveness, though, was qualified by a three-way interaction with presentation mode and time pressure, $F_{(1,\ 3298)} = 5.50$, $p = .019$. This interaction indicated that a forced delay moderated the effect of false negatives on effectiveness in the Visual mode, $B = -2.39$, $t = -2.20$, $p = .025$, such that the negative effect was present following a forced delay ($B = -1.54$, 95% CI [−2.66, −0.42]) but absent under time pressure ($B = 0.84$, 95% CI [−1.03, 2.73]). The forced delay had no corresponding influence in the Numerical mode, $B = 1.31$, $t = 1.14$, $p = .25$–where the trend was, if anything, weaker after a forced delay ($B = -1.01$, 95% CI [−2.10, 0.06]) than under time pressure ($B = -2.33$, 95% CI [−4.29, −0.36]).

Importantly, a forced delay did not moderate any of the corresponding effects of false positives on effectiveness, $ps > .30$, or justice evaluations, $ps > .34$, whether in the numerical or visual modes. Thus, the emphasis on false positives may be relatively automatic, whereas directing attention toward false negatives requires additional time and/or cognitive resources.

Study 2 replicated the influence of false positives on effectiveness observed in Study 1. So, once again, we assessed whether this effect could be indirect via approval. Approval and effectiveness were strongly positively related, $B = 0.52$, 95% CI [0.50, 0.54], $t = 57.41$, $p < .001$. Including approval in the model rendered the influence of false positives on effectiveness non-significant, $B = -0.13$, 95% CI [−0.61, 0.35], $t = -0.53$, $p = .60$. Online Supplement 2 reports further analyses of the proposed model linking the causes of approval to its justification.

3.3 Discussion

Study 2 replicated and extended our primary findings. As in Study 1, participants' evaluations of proxy offenses depended primarily on the rate at which citizens innocent of the primary wrongdoing would be convicted. This pattern arose regardless of whether information about the probative value of the proxy conduct was presented in numerical or visual form, and whether participants had limited or unlimited time to interpret the evidence. Thus, the emphasis on false positive rate appears to reflect a spontaneous cognitive process.

The likelihood of acquitting offenders did not have an observable influence on participants' decisions in Study 1. So, in Study 2, we asked whether encouraging participants to reflect might bring about greater concern for the false negative rate. In the aggregate, Study 2 revealed *some* concern for false negatives–though these effects were modest when compared to the impact of false positives. Furthermore, the false negative rate reduced approval of proxy criminal policies only under specific experimental conditions–namely, when participants had unlimited time to interpret the evidence and the evidence was presented in visual form. All in all, Study 2 suggested that false negatives can undermine support for proxy criminal statutes, but this relationship is weak and demands greater cognitive effort.

Why a visual mode of presentation would evoke concern for false negatives is unclear. Responses to the training tasks suggested that, if anything, visual presentation of the evidence *hindered* participants' efforts to interpret the data relative to the numerical summary condition. Therefore, ease of interpretation cannot easily explain the selective effect of false negatives in the visual presentation mode. Alternatively, it may be that a visual display facilitates the *valuation* of false negatives: People's representation of the value of false negatives may be more responsive to visual changes (i.e., in location, shape and hue) than to equivalent changes represented using numerals.

4 Study 3

In Studies 1 and 2, we stipulated a base rate of crime between 10% and 30%, reflecting the assumption that, in the real world, only a minority of citizens are engaged in crime. Thus, a change in the false positive rate amounts to a larger *number* (or frequency) of innocents convicted than an equivalent change in the false negative rate (amounts to in the frequency of acquitted criminals).

This raises the possibility that the differential effect of (e.g., false positive versus false negative) *rates* in fact reflects comparable effects of frequencies–undermining the seeming asymmetry between false conviction and false acquittal.[2] In Study 3, we evaluate this possibility by matching the distributions from which the *frequencies* of false positives, false negatives and true positives are drawn. A further advantage of this approach is that it would neutralize any discrepancies that may stem from the phenomenon of *diminishing sensitivity* to numbers (Friedrich et al. 1999; Stevens 1975; Tversky and Kahneman 1981). As a further measure of the relative importance of false positives versus false negatives, we record the order in which participants request to receive the information.

4.1 Methods

4.1.1 Power Analysis

We re-analyzed data from Studies 1 and 2 to calculate the partial correlation betweeαn approval and the frequency of true positives ($r = .03$, $p = .015$), false positives ($r = -.19$, $p < .001$), true negatives ($r = .10$, $p < .001$) as well as false negatives ($r = .01$, $p = .65$). A sample of 900 participants afforded 90% power to detect an effect as small as $r = .11$, setting the α level to .05.

4.1.2 Participants

941 English native speakers with at least a high school/secondary education were recruited through Prolific in September 2022 and completed the survey (454 women, age: mean = 34.9 years, SD = 11.7). Participants who failed a simple attention check were not allowed to complete the survey and their responses were not recorded.

4.1.3 Procedure

All participants read one vignette (the *Cult* scenario in the Precursor version) and were provided three pieces of information: the number of false positives, true positives, and false negatives (in a population of 1000 citizens). Participants were told that they might be provided one, two, or all three pieces of information before being asked to decide whether to approve the proxy crime, and that therefore they ought to request the most important information first. Ultimately, participants were provided all three statistics but we recorded the order in which they chose to retrieve the information as a further measure of the relative importance of false positives, true positives, and false

[2] We thank an anonymous referee for raising this objection.

negatives. The number of false positives, false negatives, and true positives were each drawn independently from a uniform distribution from 0 to 100.

4.1.4 Measures

In Study 3 we recorded three independent measures for each participant: the number of *false positives*, *false negatives*, and *true positives*. The main dependent measure was *approval* on a 7-point Likert scale. An additional dependent measure for each participant and statistic was the *order of retrieval,* ranging from 1 (first) to 3 (last).

4.2 Results

4.2.1 Effects of True Positive, False Positives, and False Negatives

Approval correlated with the frequency of true positives ($r=.20$, 95% CI [.14, .26]), and false positives ($r=-.26$, 95% CI [−.32, −.20]) positives, and weakly with false negatives ($r=-.11$, 95% CI [−.17, −.05]), all ps < .001. In a multiple regression, all three effects remained statistically significant (see Table 3).

In this same model, we conducted linear hypothesis tests to compare the magnitude of the effects of each parameter. Linear hypothesis tests ask whether model fit worsens by establishing a certain constraint: For example, in our first case, we force the coefficients of false positives and false negatives to be equal (*false positives - false negatives* = 0). This linear hypothesis test was significant, $F=9.45$, $p=.002$, indicating that the coefficients of false positives and false negatives differ significantly. The same was true of the comparison between the coefficients of true positives and false negatives: The effect of true positives was greater than the effect of false negatives (*true positives* + *false negatives* = 0), $F=4.80$, $p=.029$. The effects of true and false positives (*true positives* + *false positives* = 0), however, did not statistically differ, $F=0.95$, $p=.33$.

4.2.2 Differences in Retrieval Order

Retrieval order (1st, 2nd, or 3rd) varied across the three statistics (i.e., false positives, true positives, and false negatives) in a McNemar's test, $\chi^2=369.2$, $p<.001$. To examine differences in retrieval order by statistic, we regressed retrieval order on statistic as a fixed factor (and participant as a random effect). False positives ($M=1.48$) tended to be retrieved before true positives ($M=1.94$), $t=14.8$, and true positives before false negatives ($M=2.59$), $t=20.7$, both ps < .001.[3]

[3] An additional analysis confirmed that retrieval order was associated with the 'importance' of the information. We 'flipped' the model such that frequency was the dependent variable and approval, order and the approval × order interaction were entered as fixed factors (with random effects of content and participant). We observed a main effect of approval, $F=104.3$, $p<.001$, and an approval × order interaction, $F=5.44$, $p=.004$. The interaction indicated that the association between approval and frequency varied as a function of retrieval order: Changes in the frequency of the first and second pieces of information had a stronger effect than changes in the third piece of information (by order of retrieval); 1st vs. 3rd: $t=3.13$, $p=.005$; 2nd vs. 3rd: $t=2.40$, $p=.044$ (whereas the first and second did not differ, $t=0.83$, $p=.68$).

Table 3 Effects of false positive, true positive and false negative frequency on approval: Study 3

Effect	Estimate	*95% CI*		t	*p*
		Lower	Upper		
Intercept	3.69	3.33	4.04	20.50	.005
False positive	−1.61	−2.00	−1.22	−8.17	<.001
True positive	1.34	0.95	1.73	6.73	<.001
False negative	−0.73	−1.12	−0.33	−3.57	<.001

4.2.3 Diminishing Sensitivity as Exponential Decay

Visual inspection of our results suggested a nonlinear relationship between frequencies and approval (see Fig. 4), resembling exponential decay (Stevens 1975). Using the *minpack.lm* package, we fit three separate exponential models of the relationship between frequency and approval, where y is approval and x is the frequency (i.e., of false positives, false negatives, or true positives):

$$y = a \times e^{b \times x} + c$$

The exponential model of the false positive frequency, $y = -1.66\ e^{-0.04f} + 3.60$, provided better fit ($AIC = 3836$) than the linear model ($AIC = 3850$), $F = 15.95$, $p < .001$. The exponential model of true positive frequency, $y = 2.03\ e^{-0.04f} + 2.70$, also provided better fit ($AIC = 3865$) than the linear model ($AIC = 3873$), $F = 10.48$, $p = .001$. Meanwhile, the corresponding comparison of false negative models was not significant, $F = 0.95$, $p = .33$ (see Fig. 5).

4.3 Discussion

False positives exerted a stronger effect than did false negatives–when matched by frequency. This result replicates the primary finding of Studies 1 and 2—and demonstrates that the disparity between the influence of false positive and false negative *rates* was not due to a difference in their raw counts. Furthermore, false positives had the highest priority in participants' retrieval–such that participants tended to request information about the frequency of false positives first. Meanwhile, false negatives had the lowest priority; that is, participants tended to request the frequency of false negatives last.

Comparing the effects of false positives versus true positives revealed that these effects were (i) similar in magnitude, and (ii) subject to psychophysical numbing or diminishing sensitivity (as documented by the pattern of exponential decay; see Stevens 1975).

5 Individual Differences Analyses

In Studies 1 and 2, participants completed the following individual difference measures, which we analyze in the aggregate below:

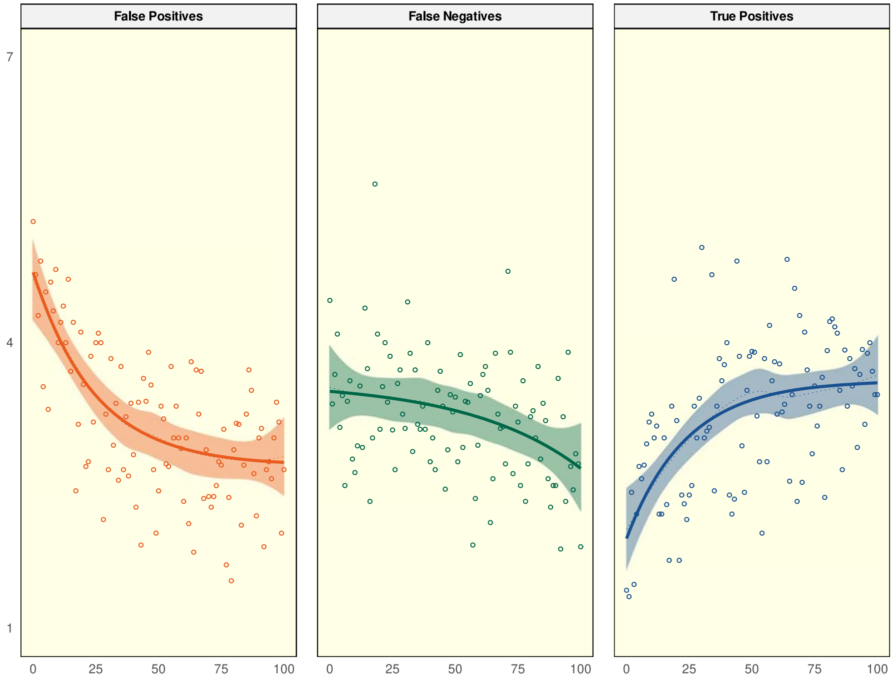


Fig. 4 Mean approval by frequency and statistic

1. *probability literacy*, the sum score of correct responses in the training task;
2. *punitive tendencies*, the six-item average (Cronbach's $\alpha = .85$) of the *Punitiveness Questionnaire* (Kemme et al. 2014);
3. *punishment theory*: a single-item, dichotomous preference for either retributivism or deterrence as the theory of punishment*;*
4. *retribution motives*: a single-item assessment of whether retributive motives guided their evaluations of proxy crimes, and
5. *deterrence motives*: a single-item assessment of whether deterrence motives guided their evaluations of proxy crimes.

The items assessing retribution and deterrence motives were adapted from work by Carlsmith and colleagues (Carlsmith et al. 2002).

Participants tended to endorse the deterrence theory of punishment (76%) over retributivism (24%) in the abstract. They also reported greater emphasis on deterrence motives than on retributivist motives ($t = 24.21$, $p < .001$, Cohen's $d = 0.87$), when asked to reflect on the motives that guided their decisions in the context of proxy crimes.

Increased punitiveness was associated positively with deterrence motives, and especially retributivist motives. Meanwhile, probabilistic literacy was negatively correlated with punitiveness and retributivist motives (see Table 4).

Attitudes toward punishment influenced participants' evaluations of proxy criminal statutes (see Online Supplement 3). As expected, punitive individuals demonstrated a more favorable attitude toward proxy crimes–which did not depend on the proxy

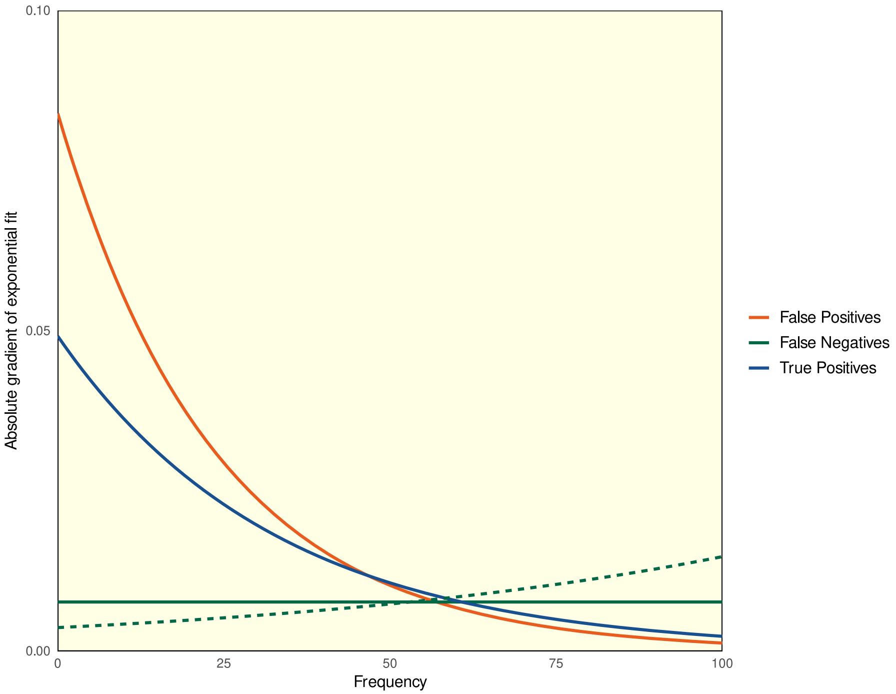


Fig. 5 Absolute gradient of the exponential fit of false positives, false negatives, and true positives. As shown, each additional false positive and true positive had a diminishing effect on approval

crimes' specificity and sensitivity. Then, when examining participants' theories of punishment, we corroborated a pattern previously observed within legal scholarship (Teichman 2017): Specifically, endorsement of deterrence theory was associated with more favorable views of proxy crime than was endorsement of retributivism.

Probability literacy, as measured on the training task, was unrelated to overall approval of proxy crimes. Higher scores on the training task were, however, tied to a greater emphasis on false positives. In other words, decisions to approve or oppose proxy criminal policies were more closely related to false positive rates among those participants who scored highly on the training task (than among those with lower scores).

6 General Discussion

To what extent is people's preference for criminal policies effectively curtailing crime constrained by their respect towards the rights of the innocent? Do people's assessment of over-inclusive criminal statutes demonstrate the operation of an intuitive Blackstone's ratio? To answer these questions, we conducted a series of experiments probing people's reactions toward a series of proxy criminal statutes. Because proxy crimes target wrongdoing indirectly, and incur in both false positives (when they result in convicting people innocent of the actual wrongdoing) and

Table 4 Individual Differences: Correlation Table ($N = 1551$)

	M (SD)	(1)	(2)	(3)	(4)
(1) Punitiveness	5.36 (1.09)	–			
(2) Retributivism	4.95 (1.39)	.44 [.40, .48]	–		
(3) Deterrence	5.98 (0.92)	.25 [.20, .29]	.11 [.06, .16]	–	
(4) Probability Literacy	3.42 (1.12)	−.14 [−.18,-.09]	−.08 [−.13, −.04]	−.05 # [−.10, .00]	–

#: non-significant ($p > .05$). All other p values below .001

false negatives (when they might lead to acquitting actual criminals), they provide a unique window into the way concerns about crime reduction and protection of the innocent guide people's evaluations of criminal policy, dissociating the impact of each.

Our studies documented a predominant and univocal commitment to the principle that it is better "that ten guilty persons escape than that one innocent suffer" (Blackstone 1765).[4] This intuitive commitment arose among individuals with diverse explicit views on the function of criminal law, and under various experimental conditions. Specifically, this effect of false positives was remarkably robust to the imposition of time pressure favoring an intuitive cognitive style, variation in the mode of presentation (i.e., whether the data were presented visually or numerically), and arose regardless of people's probability literacy and beliefs about the purpose of legal punishment.

Still, participants tended retrospectively to adduce a greater focus on the statutes' capacity to combat and deter crime. In this regard, our studies revealed a striking discrepancy in laypeople's thinking about criminal law. When introspecting about the motives driving their criminalization decisions, participants alleged both a concern with tackling crime as well as with safeguarding the innocent's rights. Taken together, these results highlight a dissociation between the factors that shape intuitive attitudes toward criminalization and the reasoning that people retrospectively offer. As such, our present studies dovetail with a growing body of research on legal decision-making (Carlsmith et al. 2002; Struchiner et al. 2020; Costa et al. 2019), documenting a recurring discrepancy between people's intuitive legal judgment and their explicit avowals.

Our results may bear on the debate between two broad camps that have dominated the theoretical landscape of criminal law. *Consequentialists* argue that new criminal offenses may be rightfully introduced as long as their benefits, primarily, their effectiveness in combating crime, outweigh their social costs. For example, the decision to approve a travel ban should rely on a calculus integrating both the ban's capacity to hinder terrorist operations and intercept the terrorists themselves, as well

[4] This is based on the assumption that we can calculate the Blackstone's ratio by dividing the observed effect of false negatives by the observed effect of false positives. Under an alternative interpretation, in which the ratio's numerator consists of true positives (rather than false negatives), the resulting ratio would be smaller. We thank an anonymous referee for suggesting this qualification.

as its detriment to well-meaning travelers. If the former exceeds the latter, there is reason to support the proxy crime—otherwise not (Teichman 2017).

In contrast, *non-consequentialists* advocate certain categorical constraints on the legitimate scope of criminalization—one of which is non-infringement on the rights of the innocent. From a non-consequentialist perspective, convicting the innocent violates a fundamental tenet of criminal law, and is therefore wrong even if doing so would come with enormous benefits for a law's expediency—and, in turn, for social welfare. Specifically, *negative retributivism* is, roughly, the claim that the state has a categorical obligation not to punish innocents nor punish the guilty more than they deserve; but it does not have a similar moral obligation to punish all offenders (Bystranowski 2017; Hoskins and Duff, 2021).

Our results provide evidence that people endorse the principles and aims of deterrence theory, yet their intuitions better align with (negative) retributivism: In the abstract, people profess to care about the expediency of a criminal policy, yet in practice their judgments are largely determined by the cost of falsely convicting the innocent and weakly by the cost of acquitting wrongdoers. This result is readily interpretable as the manifestation of negative retributivist principles; yet, it remains possible that the differential magnitude of false positive and false negative effects is the product of consequentialist reasoning—if, for example, individuals impute various indirect and downstream costs to false conviction (e.g., because of the deadweight loss of imprisonment and social stigma, or because of compromising the legitimacy and authority of the legal system) but not false acquittal.

In our interpretation of these results, we have treated the specificity-sensitivity trade-off inherent to proxy crimes as representative of the general trade-offs that permeate the entire criminal process (e.g., in the context of criminal trial). Still, strictly speaking, whether our present findings generalize beyond the specific context of proxy crimes remains to be examined in future research.

In sum, our studies offer evidence of a remarkably robust principle in laypeople's reasoning about proxy criminalization: People demonstrate much greater concern for false positives than for false negatives. This pattern emerges quickly in thought, regardless of how information about the crime is presented or of people's overt beliefs about the goal of criminal punishment. This way, our studies contribute to a growing understanding of laypeople's reasoning in the legal domain and illustrate how people intuitively manifest a negative retributivist tendency to protecting the innocent while espousing a primary commitment to mitigating crime.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s13164-023-00674-0.

Acknowledgements Piotr Bystranowski's work on Study 1 was supported by the National Science Centre, Poland (2016/23/N/HS5/00928), his work on Study 2 was supported the European Research Council (ERC) under the H2020 European Research Council research and innovation program (agreement 805498), and his work on Study 3 was supported by the Faculty of Philosophy under the Strategic Programme Excellence Initiative at Jagiellonian University. Ivar R. Hannikainen's work was supported by the Spanish Ministry of Science and Innovation (MICINN; PID2020-119791RA-I00). Participants at the Symposium on Topics in Experimental Philosophy of Law at the 2021 Conference of the European Society for Philosophy and Psychology and at the faculty meeting of the Faculty of Law of Instituto Tecnológico Autónomo de México provided useful feedback on earlier drafts.

Declarations

Conflict of Interest The authors declare no conflict of interest.

References

Alexander, L., and K.K. Ferzan. 2009. *Crime and culpability: A theory of criminal law*. Cambridge University Press.

Almagro, M., I.R. Hannikainen, and N. Villanueva. 2022. Whose Words Hurt? Contextual Determinants of Offensive Speech. *Personality and Social Psychology Bulletin* 48 (6): 937–953. https://doi.org/10.1177/01461672211026128.

Bentham, J. (1887). *Theory of legislation* (R. Hildreth, Trans.). Trübner & Company.

Blackstone, W. (1765). *Commentaries on the Laws of England*.

Bystranowski, P. 2017. Retributivism, consequentialism, and the risk of punishing the innocent: The troublesome case of proxy crimes. *Diametros* 53: 26–49.

Bystranowski, P., and M.C. Mungan. 2022. Proxy crimes. *American Criminal Law Review* 59 (1): 1–38.

Carlsmith, K.M. 2008. On justifying punishment: The discrepancy between words and actions. *Social Justice Research* 21 (2): 119–137. https://doi.org/10.1007/s11211-008-0068-x.

Carlsmith, K.M., J.M. Darley, and P.H. Robinson. 2002. Why do we punish? Deterrence and just deserts as motives for punishment. *Journal of Personality and Social Psychology* 83 (2): 284.

Costa, L.L.F., A.B.D. Esteves, R. Kreimer, N. Struchiner, and I. Hannikainen. 2019. Gender stereotypes underlie child custody decisions. *European Journal of Social Psychology* 49 (3): 548–559.

Darley, J.M., K.M. Carlsmith, and P.H. Robinson. 2000. Incapacitation and just deserts as motives for punishment. *Law and Human Behavior* 24 (6): 659–683.

De Guttry, A., F. Capone, and C. Paulussen. 2016. *Foreign fighters under international law and beyond*. Springer.

Duff, R.A. 1997. *Criminal attempts*. Oxford University Press.

Duff, R.A., L. Farmer, S. Marshall, and V. Tadros. 2007. *The trial on trial: Volume 3. Towards a normative theory of the criminal trial*. Hart Publishing.

Friedrich, J., P. Barnes, K. Chapin, I. Dawson, V. Garst, and D. Kerr. 1999. Psychophysical numbing: When lives are valued less as the lives at risk increase. *Journal of Consumer Psychology* 8 (3): 277–299.

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

Hoskins, Z. and A. Duff. 2021. Legal Punishment. In E. N. Zalta (Ed.) *The Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/entries/legal-punishment/

Husak, D. 2008. *Overcriminalization: The limits of the criminal law*. Oxford University Press.

Kemme, S., M. Hanslmaier, and C. Pfeiffer. 2014. Experience of parental corporal punishment in childhood and adolescence and its effect on punitiveness. *Journal of Family Violence* 29 (2): 129–142.

Lee, Y. 2021. Proxy crimes and Overcriminalization. *Criminal Law and Philosophy*: 1–16.

McAdams, R.H. 2005. The political economy of entrapment. *J. Crim. L. & Criminology* 96: 107.

McFatter, R.M. 1982. Purposes of punishment: Effects of utilities of criminal sanctions on perceived appropriateness. *Journal of Applied Psychology* 67 (3): 255.

Picinali, F. 2018. Can the reasonable doubt standard be justified? A reconstructed dialogue. *Canadian Journal of Law & Jurisprudence* 31 (2): Article 2.

Richard, F.D., C.F. Bond, and J.J. Stokes-Zoota. 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.

Sharp, F.C., and M.C. Otto. 1910. A study of the popular attitude towards retributive punishment. *International Journal of Ethics* 20 (3): 341–357.

Stevens, S.S. 1975. *Psychophysics: Introduction to its perceptual neural and social prospects*. New York: Wiley.

Struchiner, N., G.F.C.F. Almeida, and I.R. Hannikainen. 2020. Legal decision-making and the abstract/concrete paradox. *Cognition* 205: 104421.

Stuntz, W.J. 2001. The pathological politics of criminal law. *Michigan Law Review* 100: 505.

Teichman, D. 2017. Convicting with reasonable doubt: An evidentiary theory of criminal law. *Notre Dame Law Review* 93 (2): 757–810.

Tomlin, P. 2013. Extending the golden thread? Criminalisation and the presumption of innocence. *Journal of Political Philosophy* 21 (1): 44–66.

Tversky, A., and D. Kahneman. 1981. The framing of decisions and the psychology of choice. *Science* 21: 1–30.

Wylie, J., and A. Gantman. 2023. Doesn't everybody jaywalk? On codified rules that are seldom followed and selectively punished. *Cognition* 231: 105323.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.