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

Rong Wang[1], Yang Liu[1], & Tingting Wu[1]

[1] Nanjing Normal University

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

Instruction

Hypothesis

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

The hypothesis of experiment 4a: Experiment 4a examined whether an activated motivation to punish would lead to an increase in the legitimacy and justifiability of punishment for the phantom rule itself, the author expected social norm violations to activate punishment motivations.Experiment 4a examined the interplay between phantom rule violations alone and those committed in conjunction with a social norm violation.

The hypothesis of experiment 4b:To test whether the effect in Experiment 4a is special to phantom rules (vs other similar legal violations), authors next compared the motivated enforcement of phantom rule violations committed alone compared to those committed in tandem with a social norm violation to that same combination for prototypical legal violations. In other words, the authors sought to establish discriminant validity.

Data

The participants of 4a were collected from 207 participants using Prolific (118 females, 82 males, 4 other, 75.4% White, 17.9% People of Color, 6.7% Other or did not report, $M_{age} = 33.07$,$SD_{age} = 11.28$).

The participants of 4b were collected from 310 participants using Prolific (170 females, 133 males, 7 other, 62% White, 34% People of Color, 4% Other or did not report,$M_{age} = 33.5$,$SD_{age} = 12.2$).

Methods

Methods of 4a:A within-subjects factor (Rule Type) with two levels (phantom rule violation alone vs. social norm violation and phantom rule violation) was used to investigate whether previous social norm violations heightened judgments of the legitimacy of punishment and blameworthiness for phantom rule violations. Participants read a total of six vignettes (6 total blocks) that were paired with a face and common name, one at a time. They first saw a randomly presented block with a vignette and an image, then answered questions about punishment, blame, and morality.The authors used linear mixed-effects models that specify random intercepts for participants and stimuli.

Methods of 4b:Two within-subjects factors Scenario Type (phantom rule violation alone vs. social norm violation and phantom rule violation) and Rule Type (phantom rule vs. prototypical legal violation) were used to investigate whether previous social norm violation heightened judgments of the legitimacy of enforcement and blameworthiness for phantom rule violations and not prototypical legal violations. The procedure was the same as that used in Experiment 4a.

Results of Experiment 4a

Using linear mixed-effects models that specify random intercepts for participants and stimuli. In contrast to Experiment 3, only a few models did not meet the pre-registered ICC cut-off of 0.10 and for those models,the authoresnote the deviation and drop the stimulus random effect.

Using the collapsed variable for punishment and blamethe authorestested for differences between phantom rule violations alone compared to social norm and phantom rule violations together. Critically and as predicted, phantom rule violations paired with social norm violations were rated as more justifiable to punish compared to phantom rule violations alone, $\beta = 0.29$, $SE = 0.05$, $t(1128) = 6.27$, $p < .001$, $r = 0.18$, 95% CI [0.20, 0.38]

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

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

Results of Experiment 4b

The authoresfirst examined whether Scenario Type (Rule violation alone vs.Rule and a social norm violation) and Rule Type (Phantom rule vs.Prototypical rule), and their interaction term influenced the collapsed punishment and blame judgment index.1Results suggested that phantom rule violations lead to less punishment and blame than prototypical legal violations, $\beta = -2.75$, $SE = 0.09$, t (1773) = -31.34, $p < .001$, $r = 0.60$, 95%CI [-2.92, -2.58]. No significant effect emerged for the Scenario Type, $p = .818$, 95% CI [-0.20, 0.15]. Contrary to predictions, there was no significant interaction effect, $\beta = 0.24$, $SE = 0.13$, t (1774) = 1.82, $p = .061$, $r = 0.04$, 95% CI [-0.01, 0.49].

Following analyses in Experiment 4a,the authoresalso tested whether Scenario Type and Rule Type influence the dichotomous punishment judgment. As predicted, phantom rule violations alone were less likely to lead to selection of the punishment option (vs. the

"No punish") than prototypical legal violations, β = -3.26, SE = 0.22, z = -14.51, $p <$ .001, r = -0.67, 95% CI [-3.71, -2.83], OR = 0.04, and rules coupled with a social norm violation were more likely to lead to selection of the punishment than alone, β = 0.53, SE = 0.25, z = 2.13, p = .033, r = 0.14, 95% CI [0.05, 1.02], OR = 1.69. There was no significant interaction, p = .75, 95% CI [-0.68, 0.48]

Next,the authoresexamined whether Scenario Type and Rule Type affected the collapsed moral judgment index. As predicted, phantom rule violations themselves were rated as less morally wrong than prototypical legal violations, β = -2.67, SE = 0.08, $t(1735)$ = -32.83, $p <$ .001, r = 0.62, 95% CI [-2.83, -2.52]. There was no significant difference between the two Scenario Type conditions, p = .35, 95% CI [-0.09, 0.24].

Critically, results revealed a significant interaction, β = 0.45, SE = 0.12, $t(1736)$ = 3.83, $p <$ .001, r = 0.09, 95% CI [0.22, 0.69], such that phantom rule violations committed alone were seen as less morally wrong than those committed with a social norm violation, β = -0.53, SE = 0.09, $t(1750)$ = -6.26, $p <$ .0001 There was again no statistically significant difference between prototypical legal violations committed alone and those committed alongside a social norm violation, p = .79. Violations of phantom rules are specifically subject to motivated rule enforcement.

Procedure

Repeat ideas: According to the introduction of each part of the paper experiment, mainly focus on the data part of the paper, from the calculation of the subject's demography Analyze to other parts. The parts are following: 1. data cleaning, 2. demographics, 3. items collapsed into variable (calcualte Cronbach's α) for each dependent variable, 4. calculate the correlation of dependent variables, 5. main analyses for Experienment 4a and 4b.

All models were linear mixed models fitted in R using the lme4 package. For linear mixed models, p-values were obtained via Satterthwaite's method using the lmerTest package. Random effects structures were determined by beginning with the maximal structure and removing slopes and intercepts until the model converged. The maximal random effects structures were determined uniquely for each model and included random intercepts by sceneType and by Subject along with random slopes by each within-subject variable.

Please pay attention, considering the original paper does not start from the full model, only presents the zero model,we used more models than papers. Models with boundary (singular) fit and with strong collinearity of random effects were removed, and compared other models to select the appriate model for following analysis.

Based selected model, for experiment 4a, we examined the effect of the type of rule violation on punishment and liability scores, and the influence of situation type on moral behavior rating. For experiment 4b, based on specific selected models, we examined differences between phantom rule and prototypical legal violations for punishability, moral wrongness, frustraction and legitimacy of enforcement judgments.

Moreover, we used tables and graphs to express the results graphically. The authoresused R (Version 4.2.2; R Core Team, 2022) and the R-packages *afex* (Version 1.3.0; Singmann, Bolker, Westfall, Aust, & Ben-Shachar, 2023), *apaTables* (Version 2.0.8; Stanley, 2021), *bruceR* (Version 0.8.10; Bao, 2023), *car* (Version 3.1.2; Fox & Weisberg, 2019; Fox, Weisberg, & Price, 2022), *carData* (Version 3.0.5; Fox et al., 2022), *correlationArticle* (Makowski, Ben-Shachar, Patil, & Lüdecke, 2020), *corrplot2021* (Wei & Simko, 2021), *data.table* (Version 1.14.8; Dowle & Srinivasan, 2023), *dplyr* (Version 1.1.0; Wickham, François, Henry, Müller, & Vaughan, 2023), *effects* (Ben-Shachar, Lüdecke, & Makowski, 2020; Fox, 2003; Fox & Hong, 2009; Version 4.2.2; Fox & Weisberg, 2018), *effectsize* (Version 0.8.3; Ben-Shachar et al., 2020), *emmeans* (Version 1.8.6; Lenth, 2023), *forcats* (Version 1.0.0; Wickham, 2023), *ggplot2* (Version 3.4.2; Wickham, 2016), *gridExtra*

(Version 2.3; Auguie, 2017), *interactions* (Version 1.1.5; Long, 2019), *jtools* (Version 2.2.1; Long, 2022), *kableExtra* (Version 1.3.4; Zhu, 2021), *lattice* (Version 0.21.8; Sarkar, 2008), *lme4* (Version 1.1.33; Bates, Mächler, Bolker, & Walker, 2015), *lmerTest* (Version 3.1.3; Kuznetsova, Brockhoff, & Christensen, 2017), *lubridate* (Version 1.9.2; Grolemund & Wickham, 2011), *MASS* (Version 7.3.60; Venables & Ripley, 2002), *Matrix* (Version 1.5.4.1; Bates, Maechler, & Jagan, 2023), *MuMIn* (Version 1.47.5; Bartoń, 2023), *papaja* (Version 0.1.1; Aust & Barth, 2022), *patchwork* (Version 1.1.2; Pedersen, 2022), *performance* (Version 0.10.4; Lüdecke, Ben-Shachar, Patil, Waggoner, & Makowski, 2021), *plyr* (Wickham, 2011; Version 1.8.8; Wickham, François, et al., 2023), *psych* (Version 2.3.3; William Revelle, 2023), *purrr* (Version 1.0.1; Wickham & Henry, 2023), *readr* (Version 2.1.4; Wickham, Hester, & Bryan, 2023), *reshape2* (Version 1.4.4; Wickham, 2007), *Rmisc* (Version 1.5.1; Hope, 2022), *stringr* (Version 1.5.0; Wickham, 2022), *tibble* (Version 3.1.8; Müller & Wickham, 2022), *tidyr* (Version 1.3.0; Wickham, Vaughan, & Girlich, 2023), *tidyverse* (Version 2.0.0; Wickham et al., 2019), *tinylabels* (Version 0.2.3; Barth, 2022), and *Unicode* (Version 15.0.0.1; Hornik, 2022) for all our analyses.

Results of Experienment 4a

Data were collected from 207 participants using Prolific (118 females, 82 males, 7 other, 156 White, 37 People of color, 14 Other or did not report, M_{age} = 33.07 years, SD_{age} = 11.28).

The primary goal of experiment 4a was to assess differences between the punishability of phantom rule violations when they are committed in isolation or in conjunction with a more morally consequential (i.e., social norm) violation. Firstly, Two punishment items were assessed, and based on internal reliability, a single punishment judgment variable was created by collasping across the two items(Cronbach's $\alpha = 0.85$). Second, a single Moral judgment variable was created by collasping across the two items(Cronbach's $\alpha = 0.78$). Correlations between the dependent variables are included in the Figure 1 and Table 1.

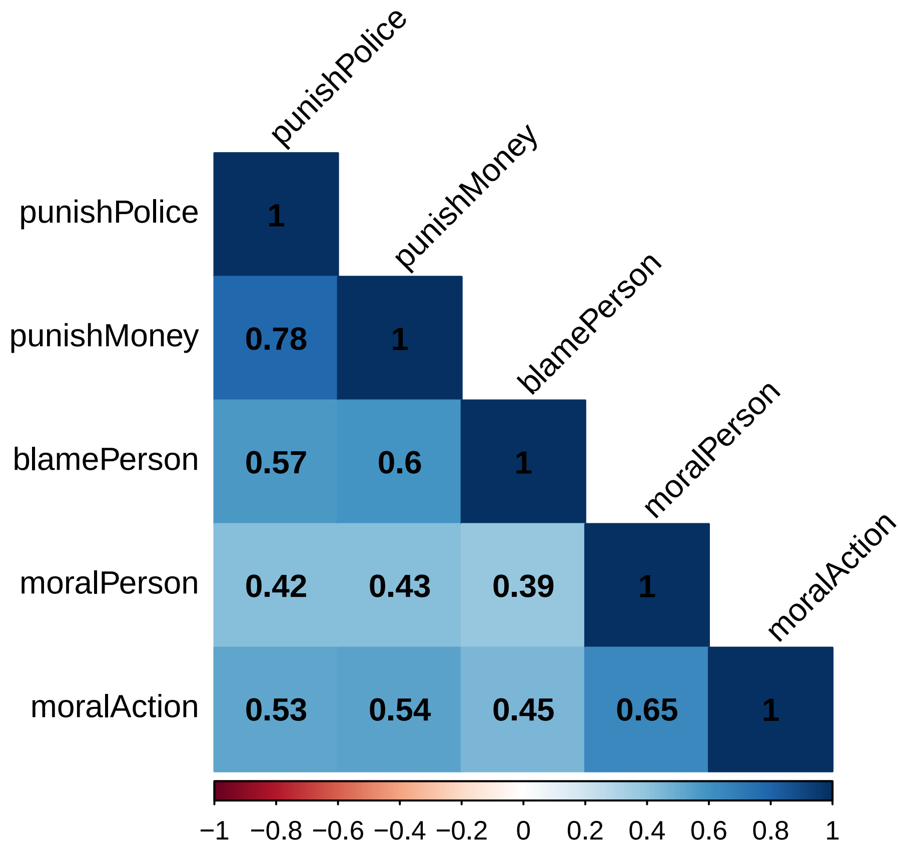

Figure 1. Correlation matrix for dependent variables.

Table 1

Corrlation matrix for dependent variables

	punishPolice	punishMoney	blamePerson	moralPerson	moralAction
punishPolice	1	<0.001	<0.001	<0.001	<0.001
punishMoney	0.784	1	<0.001	<0.001	<0.001
blamePerson	0.571	0.597	1	<0.001	<0.001
moralPerson	0.42	0.428	0.386	1	<0.001
moralAction	0.526	0.537	0.45	0.646	1

Note. The lower is correlations among variables, and the upper is the p-value.

Table 2

Model Comparison

Name	ICC	RMSE	Sigma	Performance_Score
modE2.1	0.331	1.374	1.468	0.995
modE2.2	0.268	1.437	1.528	0.395
modE2.05	0.054	1.739	1.744	0.125

Note. see Appendix A

First, the punishment and blame. Using the collapsed variable for punishment and blame we tested for differences between phantom rule violations alone compared to social norm and phantom rule violations together. Please pay attention, considering the original paper does not start from the full model, only presents the zero model.We using more models than papers, especially, constructed with the collapsed mean of punishment and blame score as the dependent variable, the type of rule violation as the independent variable (fixed factor), and the type of subjects and stories as the random factor. And the comparison results of models please see table 2. And we chose the model contained the type of subjects and stories as the random factors (i.e., modE2.1).

Similar with the results of this article, phantom rule violations paired with social norm violations were rated as more justifiable to punish compared to phantom rule violations alone, $\beta = 0.58$, $SE = 0.09$, t(1,116.80)= 6.55, p = <0.001, 95%CI = 0.41 and 0.76.

We next examined whether the phantom rule violation alone compared to social norm and phantom rule violations together influenced judgments of the morality of the particular phantom rule violation and of the person.Models with boundary (singular) fit were removed, the comparison results of other models please see table 3. And we chose the model contained the type of subjects and stories as the random factors (i.e., modE2.3).

Table 3

Model Comparison

Name	ICC	RMSE	Sigma	Performance_Score
modE2.3	0.388	1.340	1.431	1.000
modE2.36	0.235	1.493	1.582	0.307
modE2.37	0.133	1.697	1.701	0.000

Note. see Appendix B

Results revealed that when a phantom rule was committed following a social norm violation, the phantom rule violation itself was seen as more morally wrong compared to when a phantom rule violation was committed alone, $\beta = 0.40$, $SE = 0.09$, $t(1{,}115.49) = 4.58$, $p = <0.001$, $95\% CI = [0.23$ and $0.57]$.

This was also true for the moral character variable,models with boundary (singular) fit were removed, the comparison results of models please see table 4. Considering The collinearity of random effects is too strong in modE2.42 and modE2.41 (r = -0.91, r = -0.91), we chose the model contained the type of stories and subjects as the random factors, and scenariosDO as by-subject slopes (i.e., modE2.43). Violators of phantom rules paired with social norm violations were seen as more morally bad than violators of phantom rules alone, $\beta = 1.03$, $SE = 0.09$, $t(198.13) = 11.63$, $p = <0.001$, $95\% CI = 1.30$ and 1.43.

Results of Experienment 4b

Data were collected from 310 participants using Prolific (170 females, 133 males, 7 other, 193 White, 105 People of color, 12 Other or did not report, $M_{age} = 33.49$ years, $SD_{age} = 12.19$).

Two within-subjects factors Scenario Type (phantom rule violation alone vs. social norm violation and phantom rule violation) and Rule Type (phantom rule vs. prototypical

Table 4

Model Comparison

Name	ICC	RMSE	Sigma	Performance_Score
modE2.42	0.317	1.263	1.347	0.914
modE2.41	0.344	1.215	1.319	0.795
modE2.43	0.295	1.262	1.360	0.547
modE2.4	0.274	1.298	1.381	0.511
modE2.45	0.253	1.302	1.396	0.487
modE2.46	0.235	1.334	1.413	0.455
modE2.44	0.074	1.558	1.565	0.054
modE2.47	0.034	1.587	1.591	0.015

Note. see Appendix C

legal violation) were used to investigate whether previous social norm violation heightened judgments of legitimacy of enforcement and blameworthiness for phantom rule violations and not prototypical legal violations. Firstly, We used similar punishment items from Experiment 4a. Four punishment items were assessed, and based on internal reliability, a single punishment judgment variable was created by collasping across the four items(Cronbach's $\alpha = 0.93$). Second, a single Moral judgment variable was created by collasping across the three items(Cronbach's $\alpha = 0.92$). Furthermore, there are five items used to assess judgments about the legitimacy and fairness of phantom rules (Cronbach's $\alpha = 0.88$). Lastly, three items to assess the affective quality of phantom rule violation enforcement.These items had high internal reliability and were collapsed into a single frustration index (Cronbach's $\alpha = 0.86$).

First we examined whether Scenario Type (Rule violation alone vs.Rule and a social norm violation) and Rule Type (Phantom rule vs.Prototypical rule), and their interaction

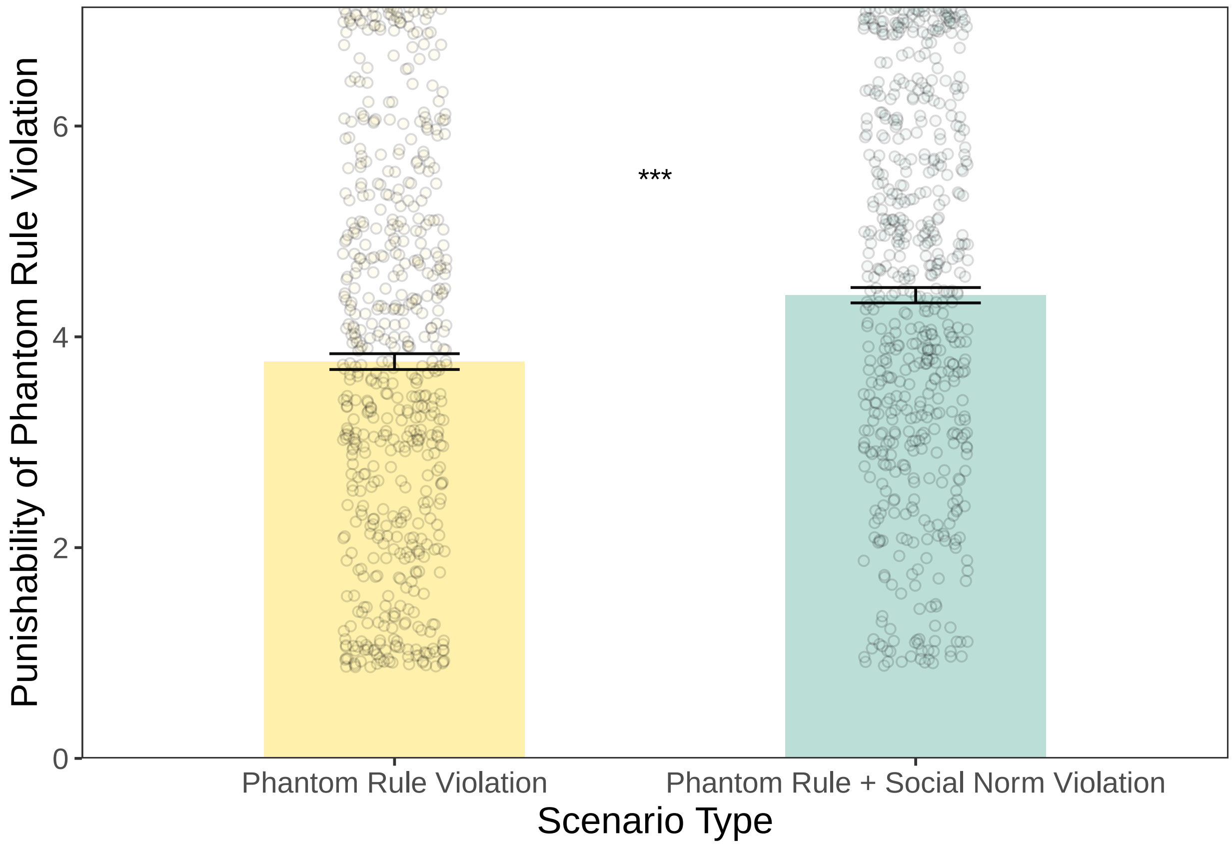

Figure 2. Punishment differences for phantom rule and social norm violations compared to phantom rule violations alone.

term influenced the collapsed punishment and blame judgment index. And the comparison results of models please see table 5. Based on selected modE5.115 (random-effects terms included intercepts for all random effects, Type as by-subject slopes, and Condition as by-sceneType intercepts), results suggested that phantom rule violations lead to less punishment and blame than prototypical legal violations(see Figure 4),β = 2.52, SE = 0.18, t(6.71)= 14.15, p = <0.001, 95%CI = [-0.41 and 0.97]. No significant effect emerged for the Scenario Type,F =3.76, p = 0.05. Contrary to predictions(similar with origin article), there was no significant interaction effect,F =2.89, p = 0.09.

```
## Random effect variances not available. Returned R2 does not account for random effect
```

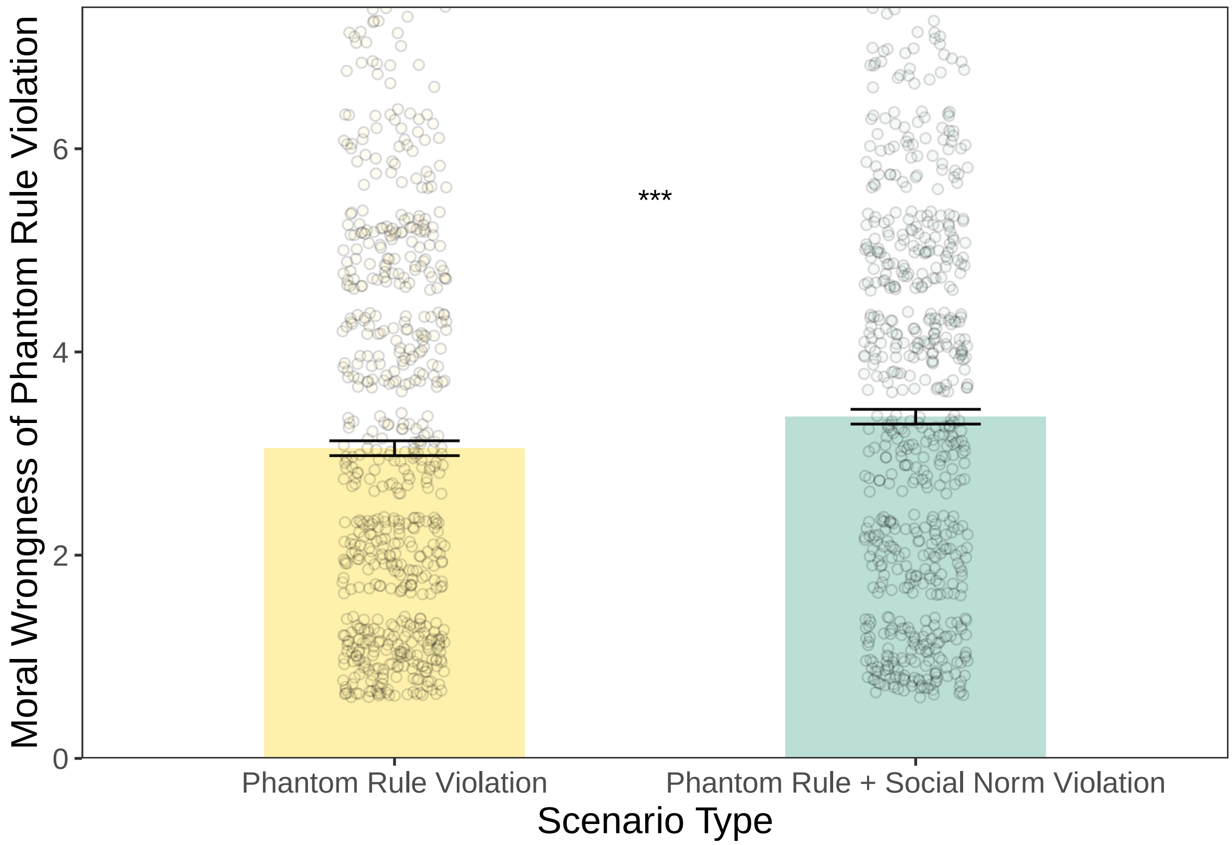

Figure 3. Moral differences for phantom rule and social norm violations compared to phantom rule violations alone.

Following analyses in Experiment 4a, we also tested whether Scenario Type and Rule Type influence the dichotomous punishment judgment.And the comparison results of models please see table 6. Based on selected modE5.28 (random-effects terms included intercepts and Type and Condition as slopes for all random effects), as predicted, phantom rule violations alone were less likely to lead to selection of the punishment option (vs. the "No punish") than prototypical legal violations(see Figure 5), $\beta = 2.21$, $SE = 0.17$, $t(7.48) = 13.32$, $p = <0.001$, and rules coupled with a social norm violation were more likely to lead to selection of the punishment than alone, $F = 8.40$, $p = 0.03$. There was significant interaction,$F = 15.60$, $p = 0.00$(differenciate with origin results). Moreover, Simple effect analysis discovered that, when only one kind of rule was broken, subjects rated the violation of only the phant om rule as less morally wrong than the violation of only the legal rule, β = -2.21,SE = 0.17, t(7.49)= -13.30, p = 0.00. When both rules were violated, participants also rated the violator of the ghost rule as less morally wrong than the violator of the legal rule,β = -2.67,SE = 0.16, t(6.93)= -16.38, p = 0.00. Only transgressions were not statistically significant from both violations of social norms and transgressions, p = 0.93.

Next, we examined whether Scenario Type and Rule Type affected the collapsed moral judgment index. See table 7 for model comparison, modE5.314 was selected (random-effects terms included intercepts and Condition as slopes for all random effects). As predicted, phantom rule violations themselves were rated as less morally wrong than prototypical legal violations(see Figure 6),β = 0.98, SE = 0.11, t(8.17)= 9.26, p = <0.001. There was no significant difference between the two Scenario Type conditions, F =0.06, p = 0.81.

Critically, results revealed also no significant interaction (differentiate with origin results),F =1.35, p = 0.25.

Finally, we also asked participants about how frustrated they would feel if they got caught for violating one of these rules. We used the same method of constructing and choosing the model as before. ModE5.414 was chosen as the best model. The model's random-effects terms included intercepts and Condition as slopes for all random effects. Full model specifications, including final random effects structures, are reported in table 8. As predicted, participants rated phantom rule violations as more frustrating to get caught for compared to prototypical legal violations(see Figure 7),β = -1.69, SE = 0.16, t(10.31)= -10.35, p = <0.001. While, there is also significant effect emerged for Scenario Type (differentiate with origin results),F =5.04, p = 0.02. A significant interaction emerged, F =5.60, p = 0.02. The results of the simple effect test found that when only one rule was violated, the punishment for violating only the ghost rule was higher than that for violating only the legal rule, indicating that the punishment was more frustrating, β =

Table 5

Model Comparison

Name	ICC	RMSE	Sigma	Performance_Score
modE5.115	0.271	1.159	1.226	0.773
modE5.116	0.262	1.171	1.234	0.634
modE5.120	0.255	1.172	1.237	0.496
modE5.122	0.246	1.183	1.245	0.454
modE5.125	0.144	1.259	1.316	0.315
modE5.117	0.108	1.352	1.357	0.103
modE5.127	0.088	1.365	1.368	0.080

Note. see Appendix D

1.99,SE = 0.16, t(9.58)= 12.46, p = 0.00.Only breaking the ghost rules is rated higher and the punishment more frustrating than breaking the social and ghost rules, β = -0.30,SE = 0.09, t(1,538.76)= -3.24, p = 0.01.

Table 6

Model Comparison

Name	ICC	RMSE	Sigma	Performance_Score
modE5.28	0.434	1.064	1.151	0.668
modE5.29	0.427	1.078	1.160	0.612
modE5.214	0.420	1.083	1.165	0.521
modE5.25	NA	1.127	1.194	0.435
modE5.213	0.426	1.071	1.157	0.368
modE5.210	0.404	1.108	1.183	0.288
modE5.215	0.397	1.116	1.190	0.281
modE5.211	0.393	1.128	1.195	0.266
modE5.216	0.386	1.134	1.200	0.262
modE5.217	0.225	1.342	1.346	0.047
modE5.227	0.215	1.350	1.353	0.038
modE5.225	0.124	1.347	1.404	0.022

Note. see Appendix E

Discussion

Most of the results were successful, and success here means that although the data were not identical, the overall trend and the general range of the data were consistent. For example, for the results of experiment 4a, similar with the results of this article, phantom rule violations paired with social norm violations were rated as more justifiable to punish compared to phantom rule violations alone, $\beta = 0.58$, $SE = 0.09$, t(1,116.80)= 6.55, $p =$ <0.001, 95%CI = 0.41 and 0.76. The origin results of article are following:Critically and as predicted, phantom rule violations paired with social norm violations were rated as more justifiable to punish compared to phantom rule violations alone, $\beta = 0.29$,$SE = 0.05$,

Table 7

Model Comparison

Name	ICC	RMSE	Sigma	Performance_Score
modE5.314	0.355	0.689	0.755	0.896
modE5.324	0.274	0.756	0.799	0.394
modE5.316	0.285	0.750	0.794	0.391
modE5.327	0.173	0.805	0.846	0.304
modE5.317	0.102	0.886	0.889	0.022
modE5.325	0.088	0.892	0.895	0.013

Note. see Appendix F

Table 8

Model Comparison

Name	ICC	RMSE	Sigma	Performance_Score
modE5.414	0.480	1.101	1.239	0.994
modE5.413	0.485	1.093	1.234	0.646
modE5.419	0.477	1.102	1.242	0.614
modE5.421	0.473	1.110	1.247	0.610
modE5.415	0.403	1.239	1.334	0.497
modE5.416	0.401	1.243	1.336	0.493
modE5.424	0.395	1.250	1.342	0.488
modE5.427	0.333	1.311	1.403	0.436
modE5.417	0.062	1.666	1.671	0.008
modE5.425	0.054	1.673	1.677	0.003

Note. see Appendix G

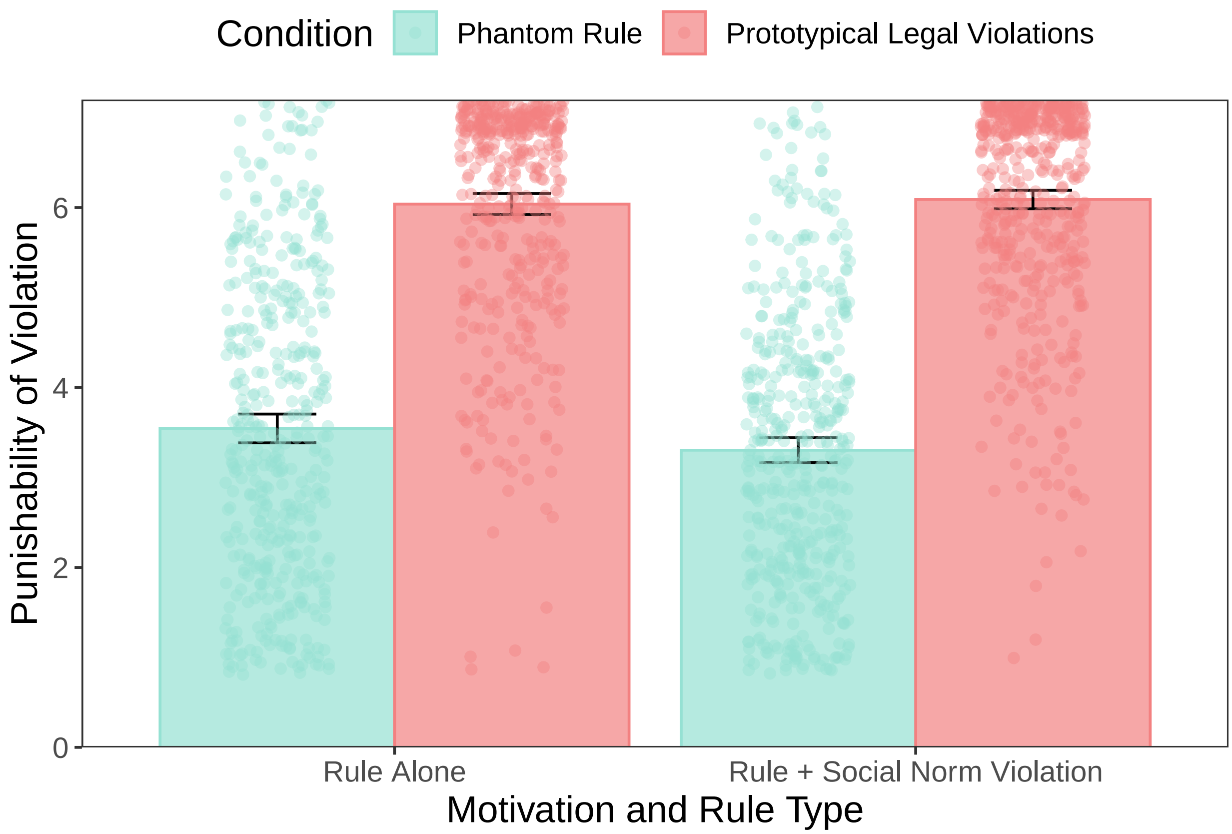

Figure 4. Differences between phantom rule and prototypical legal violations for punishability of enforcement judgments.

$t(1128) = 6.27$, $p < .001$, 95% $CI[0.20, 0.38]$.

From our perspectives, these small differences may be caused by following reasons: Firstly, considering the original paper does not start from the full model, only presents the zero model,we used more models than papers. Secondly, the differences of the contract methods. The origin article mainly made them through creating sum to zero contrasts, rather than the default method (i.e., treatment coding), which we used. In simple terms, the former reports the real beta needs *2, while the latter outputs the real Beta.

There are also some strange results we can't unnderstant. For example, the results of experiment 4b, based on selected modE5.115, results suggested that phantom rule violations lead to less punishment and blame than prototypical legal violations(see Figure

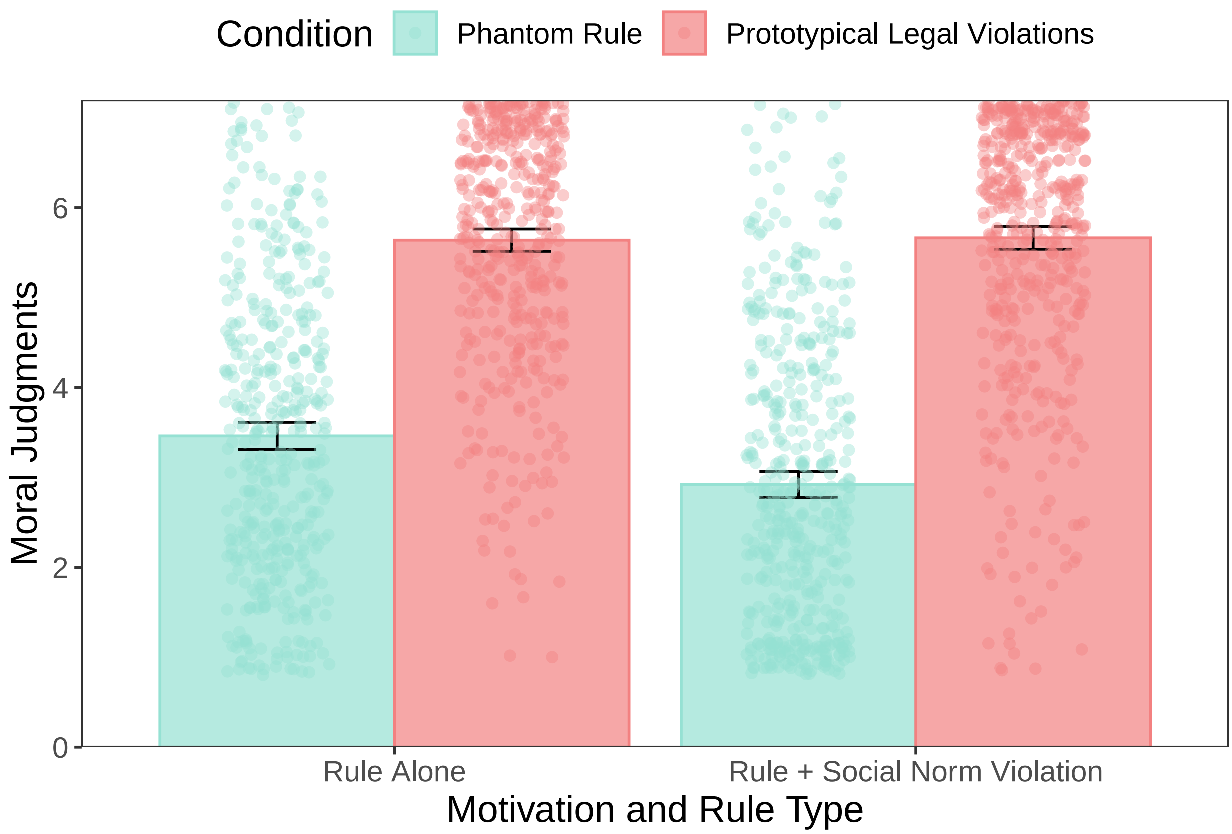

Figure 5. Differences between phantom rule and prototypical legal violations for moral wrongness of enforcement judgments.

4),β = 2.52, SE = 0.18, t(6.71)= 14.15, p = <0.001, 95%CI = [-0.41 and 0.97]. There is a strange degree of freedom for t, which is 6.71. And the original version is 1773.

Additionally, when we used the papaja to generate the dpf version with APA format, there are something unconvinient, for example, the warning of codes can't be ignored. After tring the method papaja provided (i.e., echo = FALSE, message = FALSE), and the method R provided (i.e., options(warn = -1)), this problem still didn't be resovled.

In general, it is convenient to use R for data analysis and export pdf files in APA format directly with papaja.

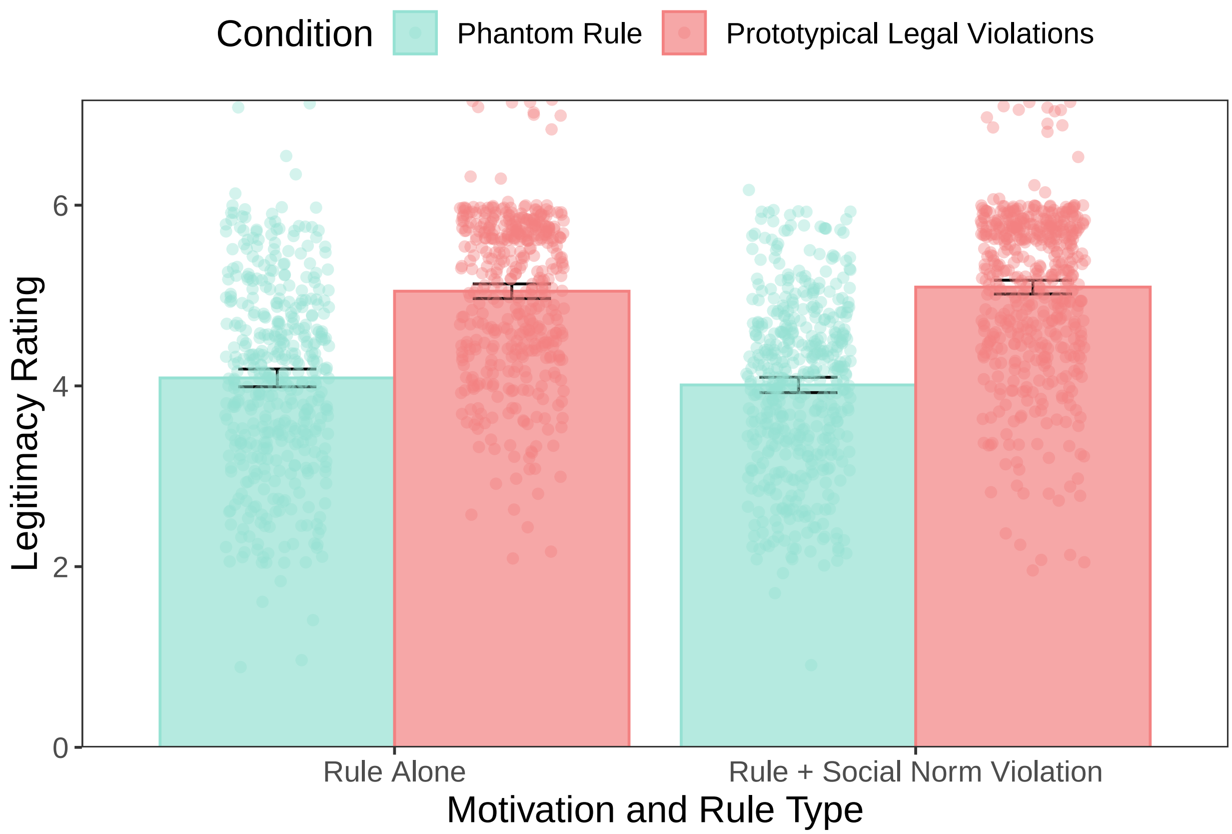

Figure 6. Differences between phantom rule and prototypical legal violations for legitimacy of enforcement judgments.

Division of Labor

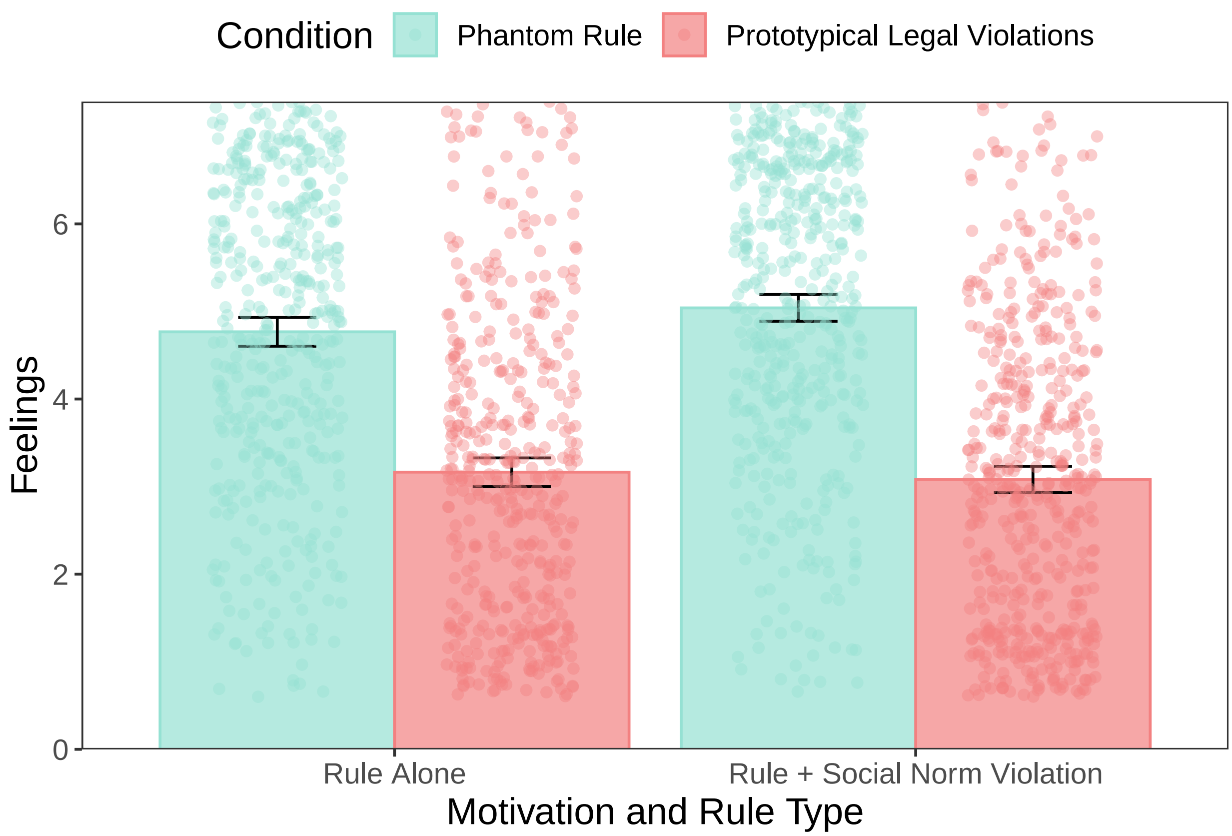

Figure 7. Differences between phantom rule and prototypical legal violations for frustration of enforcement judgments.

Table 9

Division of Labor

Assinment	Rong Wang	Yang Liu	Tingting Wu
Code	20%: library packages,data cleaning,descriaption analysis	60%: Generalised Linear Mixed Models,Model comparison	20%: Figure
Document	60%: Results of Experienment 4a and 4b, Discussion, Division of Labor	20%: Procedure	20%: Instruction
Report	20%	20%	60%

References

Auguie, B. (2017). *gridExtra: Miscellaneous functions for "grid" graphics.* Retrieved from https://CRAN.R-project.org/package=gridExtra

Aust, F., & Barth, M. (2022). *papaja: Prepare reproducible APA journal articles with R Markdown.* Retrieved from https://github.com/crsh/papaja

Bao, H.-W.-S. (2023). *bruceR: Broadly useful convenient and efficient r functions.* Retrieved from https://CRAN.R-project.org/package=bruceR

Barth, M. (2022). *tinylabels: Lightweight variable labels.* Retrieved from https://cran.r-project.org/package=tinylabels

Bartoń, K. (2023). *MuMIn: Multi-model inference.* Retrieved from https://CRAN.R-project.org/package=MuMIn

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bates, D., Maechler, M., & Jagan, M. (2023). *Matrix: Sparse and dense matrix classes and methods.* Retrieved from https://CRAN.R-project.org/package=Matrix

Ben-Shachar, M. S., Lüdecke, D., & Makowski, D. (2020). effectsize: Estimation of effect size indices and standardized parameters. *Journal of Open Source Software*, *5*(56), 2815. https://doi.org/10.21105/joss.02815

Dowle, M., & Srinivasan, A. (2023). *Data.table: Extension of ‘data.frame‘.* Retrieved from https://CRAN.R-project.org/package=data.table

Fox, J. (2003). Effect displays in R for generalised linear models. *Journal of Statistical Software*, *8*(15), 1–27. https://doi.org/10.18637/jss.v008.i15

Fox, J., & Hong, J. (2009). Effect displays in R for multinomial and proportional-odds logit models: Extensions to the effects package. *Journal of Statistical Software*, *32*(1), 1–24. https://doi.org/10.18637/jss.v032.i01

Fox, J., & Weisberg, S. (2018). Visualizing fit and lack of fit in complex regression models with predictor effect plots and partial residuals. *Journal of Statistical Software*, *87*(9), 1–27. https://doi.org/10.18637/jss.v087.i09

Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (Third). Thousand Oaks CA: Sage. Retrieved from https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Fox, J., Weisberg, S., & Price, B. (2022). *carData: Companion to applied regression data sets*. Retrieved from https://CRAN.R-project.org/package=carData

Grolemund, G., & Wickham, H. (2011). Dates and times made easy with lubridate. *Journal of Statistical Software*, *40*(3), 1–25. Retrieved from https://www.jstatsoft.org/v40/i03/

Hope, R. M. (2022). *Rmisc: Ryan miscellaneous*. Retrieved from https://CRAN.R-project.org/package=Rmisc

Hornik, K. (2022). *Unicode: Unicode data and utilities*. Retrieved from https://CRAN.R-project.org/package=Unicode

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

Lenth, R. V. (2023). *Emmeans: Estimated marginal means, aka least-squares means*. Retrieved from https://CRAN.R-project.org/package=emmeans

Long, J. A. (2019). *Interactions: Comprehensive, user-friendly toolkit for probing interactions*. Retrieved from https://cran.r-project.org/package=interactions

Long, J. A. (2022). *Jtools: Analysis and presentation of social scientific data*. Retrieved from https://cran.r-project.org/package=jtools

Lüdecke, D., Ben-Shachar, M. S., Patil, I., Waggoner, P., & Makowski, D. (2021). performance: An R package for assessment, comparison and testing of statistical models. *Journal of Open Source Software*, *6*(60), 3139. https://doi.org/10.21105/joss.03139

Makowski, D., Ben-Shachar, M. S., Patil, I., & Lüdecke, D. (2020). Methods and algorithms for correlation analysis in R. *Journal of Open Source Software*, *5*(51), 2306. https://doi.org/10.21105/joss.02306

Müller, K., & Wickham, H. (2022). *Tibble: Simple data frames*. Retrieved from https://CRAN.R-project.org/package=tibble

Pedersen, T. L. (2022). *Patchwork: The composer of plots*. Retrieved from https://CRAN.R-project.org/package=patchwork

R Core Team. (2022). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Sarkar, D. (2008). *Lattice: Multivariate data visualization with r*. New York: Springer. Retrieved from http://lmdvr.r-forge.r-project.org

Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. S. (2023). *Afex: Analysis of factorial experiments*. Retrieved from https://CRAN.R-project.org/package=afex

Stanley, D. (2021). *apaTables: Create american psychological association (APA) style tables*. Retrieved from https://CRAN.R-project.org/package=apaTables

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with s* (Fourth). New York: Springer. Retrieved from https://www.stats.ox.ac.uk/pub/MASS4/

Wei, T., & Simko, V. (2021). *R package 'corrplot': Visualization of a correlation matrix*. Retrieved from https://github.com/taiyun/corrplot

Wickham, H. (2007). Reshaping data with the reshape package. *Journal of Statistical Software*, *21*(12), 1–20. Retrieved from http://www.jstatsoft.org/v21/i12/

Wickham, H. (2011). The split-apply-combine strategy for data analysis. *Journal of Statistical Software*, *40*(1), 1–29. Retrieved from https://www.jstatsoft.org/v40/i01/

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. Retrieved from https://ggplot2.tidyverse.org

Wickham, H. (2022). *Stringr: Simple, consistent wrappers for common string operations.* Retrieved from https://CRAN.R-project.org/package=stringr

Wickham, H. (2023). *Forcats: Tools for working with categorical variables (factors).* Retrieved from https://CRAN.R-project.org/package=forcats

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., ... Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). *Dplyr: A grammar of data manipulation.* Retrieved from https://CRAN.R-project.org/package=dplyr

Wickham, H., & Henry, L. (2023). *Purrr: Functional programming tools.* Retrieved from https://CRAN.R-project.org/package=purrr

Wickham, H., Hester, J., & Bryan, J. (2023). *Readr: Read rectangular text data.* Retrieved from https://CRAN.R-project.org/package=readr

Wickham, H., Vaughan, D., & Girlich, M. (2023). *Tidyr: Tidy messy data.* Retrieved from https://CRAN.R-project.org/package=tidyr

William Revelle. (2023). *Psych: Procedures for psychological, psychometric, and personality research.* Evanston, Illinois: Northwestern University. Retrieved from https://CRAN.R-project.org/package=psych

Zhu, H. (2021). *kableExtra: Construct complex table with 'kable' and pipe syntax.* Retrieved from https://CRAN.R-project.org/package=kableExtra

Appendix

Appendix A

modE2.05 <- lmer(pandb ~ scenariosDO + (1|sceneType), data = rePD4a.Long); modE2.1 <- lmer(pandb ~ scenariosDO + (1|sceneType) + (1|Subj), data = rePD4a.Long); modE2.2 <- lmer(pandb ~ scenariosDO + (1|Subj),data = rePD4a.Long); modE2.05 did not meet the pre-registered ICC cut-off of 0.10.

Appendix B

modE2.3 <- lmer(moralAction ~ scenariosDO + (1|sceneType) + (1|Subj), data = rePD4a.Long); modE2.32 <- lmer(moralAction ~ scenariosDO + (1+scenariosDO|sceneType) + (1|Subj), data = rePD4a.Long); modE2.36 <- lmer(moralAction ~ scenariosDO + (1|Subj), data = rePD4a.Long); modE2.37 <- lmer(moralAction ~ scenariosDO + (1|sceneType), data = rePD4a.Long);

Appendix C

modE2.4 <- lmer(moralPerson ~ scenariosDO + (1|sceneType) + (1|Subj), data = rePD4a.Long); modE2.41 <- lmer(moralPerson ~ scenariosDO + (1+scenariosDO|sceneType) + (1+scenariosDO|Subj), data = rePD4a.Long); modE2.42 <- lmer(moralPerson ~ scenariosDO + (1+scenariosDO|sceneType) + (1|Subj), data = rePD4a.Long); modE2.43 <- lmer(moralPerson~ scenariosDO + (1|sceneType) + (1+scenariosDO|Subj), data = rePD4a.Long); modE2.44 <- lmer(moralPerson ~ scenariosDO + (1+scenariosDO|sceneType), data = rePD4a.Long); modE2.45 <- lmer(moralPerson~ scenariosDO + (1+scenariosDO|Subj), data = rePD4a.Long); modE2.46 <- lmer(moralPerson ~ scenariosDO + (1|Subj), data = rePD4a.Long); modE2.47 <- lmer(moralPerson~ scenariosDO + (1|sceneType), data = rePD4a.Long); The collinearity of random effects is too strong in modE2.42 and modE2.41

Appendix D

modE5.115<- lmer(pandb ~ ConditionType + (1+Condition|sceneType) + (1+Type|Subj), data = rePD4b.Long); modE5.116<- lmer(pandb ~ ConditionType + (1+Condition|sceneType) + (1|Subj), data = rePD4b.Long); modE5.117 <- lmer(pandb ~ ConditionType + (1+Condition|sceneType), data = rePD4b.Long); modE5.120<- lmer(pandb ~ ConditionType + (1+Type|sceneType) + (1+Type|Subj), data = rePD4b.Long); modE5.122<- lmer(pandb ~ ConditionType + (1+Type|sceneType) + (1|Subj), data = rePD4b.Long); modE5.127 <- lmer(pandb ~ ConditionType + (1|sceneType) , data = rePD4b.Long); modE5.125 <- lmer(pandb ~ Condition*Type + (1|Subj) , data = rePD4b.Long);

Appendix E

modE5.25<- lmer(moralRatings~ ConditionType + (1+ConditionType|sceneType) + (1|Subj), data = rePD4b.Long); modE5.28 <- lmer(moralRatings ~ ConditionType + (1+Condition+Type|sceneType) + (1+Condition+Type|Subj), data = rePD4b.Long); modE5.29 <- lmer(moralRatings~ ConditionType + (1+Condition+Type|sceneType) + (1+Condition|Subj), data = rePD4b.Long); modE5.210 <- lmer(moralRatings ~ ConditionType + (1+Condition+Type|sceneType) + (1+Type|Subj), data = rePD4b.Long); modE5.211 <- lmer(moralRatings ~ ConditionType + (1+Condition+Type|sceneType) + (1|Subj), data = rePD4b.Long); modE5.213 <- lmer(moralRatings~ ConditionType + (1+Condition|sceneType) + (1+Condition+Type|Subj), data = rePD4b.Long); modE5.214<- lmer(moralRatings~ ConditionType + (1+Condition|sceneType) + (1+Condition|Subj), data = rePD4b.Long); modE5.215<- lmer(moralRatings~ ConditionType + (1+Condition|sceneType) + (1+Type|Subj), data = rePD4b.Long); modE5.216<- lmer(moralRatings~ ConditionType + (1+Condition|sceneType) + (1|Subj), data = rePD4b.Long); modE5.217 <- lmer(moralRatings~ ConditionType +

*(1+Condition|sceneType), data = rePD4b.Long); modE5.227 <- lmer(moralRatings~ Condition*Type + (1|sceneType) , data = rePD4b.Long); modE5.225 <- lmer(moralRatings~ Condition*Type + (1|Subj) , data = rePD4b.Long);

Appendix F

modE5.314<- lmer(legitRatings~ Condition*Type + (1+Condition|sceneType) + (1+Condition|Subj), data = rePD4b.Long); modE5.316<- lmer(legitRatings~ Condition*Type + (1+Condition|sceneType) + (1|Subj), data = rePD4b.Long); modE5.317 <- lmer(legitRatings~ Condition*Type + (1+Condition|sceneType), data = rePD4b.Long); modE5.324 <- lmer(legitRatings~ Condition*Type + (1|sceneType) + (1|Subj), data = rePD4b.Long); modE5.325 <- lmer(legitRatings~ Condition*Type + (1|sceneType) , data = rePD4b.Long); modE5.327 <- lmer(legitRatings~ Condition*Type + (1|Subj) , data = rePD4b.Long);

Appendix G

modE5.45<- lmer(feelRatings~ Condition*Type + (1+Condition*Type|sceneType) + (1|Subj), data = rePD4b.Long); modE5.413 <- lmer(feelRatings~ Condition*Type + (1+Condition|sceneType) + (1+Condition+Type|Subj), data = rePD4b.Long); modE5.414<- lmer(feelRatings~ Condition*Type + (1+Condition|sceneType) + (1+Condition|Subj), data = rePD4b.Long); modE5.415<- lmer(feelRatings~ Condition*Type + (1+Condition|sceneType) + (1+Type|Subj), data = rePD4b.Long); modE5.416<- lmer(feelRatings~ Condition*Type + (1+Condition|sceneType) + (1|Subj), data = rePD4b.Long); modE5.417 <- lmer(feelRatings~ Condition*Type + (1+Condition|sceneType), data = rePD4b.Long); modE5.419<- lmer(feelRatings~ Condition*Type + (1+Type|sceneType) + (1+Condition+Type|Subj), data = rePD4b.Long); modE5.421 <- lmer(feelRatings~ Condition*Type + (1+Type|sceneType) + (1+Condition|Subj), data = rePD4b.Long); modE5.424 <- lmer(feelRatings~

*Condition*Type + (1|sceneType) + (1|Subj), data = rePD4b.Long); modE5.425 <-
lmer(feelRatings ~ Condition*Type + (1|sceneType) , data = rePD4b.Long); modE5.427 <-
lmer(feelRatings~ Condition*Type + (1|Subj) , data = rePD4b.Long);