Method and Results of the Pilot Study

for the Main Paper

Comparing and Integrating Theories of Law Obedience:

How deterrence, legitimacy, and lack of self-control influence low-level criminal behaviors

Berenike Waubert de Puiseau[1], Andreas Glöckner[2,3], & Emanuel V. Towfigh[4,3]

[1]Heinrich-Heine-University, Düsseldorf

[2]University of Hagen

[3]Max Planck Institute for Research on Collective Goods, Bonn

[4] EBS University Law School, Wiesbaden

Method

Participants and Design. Fifty-two local residents from the city of Bonn, Germany, (mainly students; 61.5% female; M_{age} = 25.2, *SD* = 7.1) were recruited from the subject pool of the MPI Decision Lab through an online recruiting system (Greiner, 2004). The study used a 6 (scenarios of mid-range criminal activities) x 2 (detection probability: low vs. high) x 2 (financial benefit: low vs. high) nested mixed-factorial design with scenarios as repeated measurement factor. Detection probability and financial benefit were manipulated in two of the six scenarios as between-subjects factor. Participants were randomly assigned to the four experimental conditions and completed the experiment following an unrelated study. Participants were paid a fixed fee of 12 Euros (16.80 USD) for their participation in the one-hour session. For one person, data on the predictor scales were partially missing, and the person was excluded from the respective analyses.

Materials. To measure the participants' tendency to break the law, we constructed six scenarios in which a person needs to decide either to break or obey the law. The scenarios were designed to tempt participants to disobey the law. Topics were free-riding on the train, illegal downloading of music, speeding, insurance fraud, illegal parking, and littering in a nature reserve. Financial benefit was manipulated in the free-riding scenario; detection probability was manipulated in the insurance fraud scenario. These two vignettes, including the manipulations, read as follows (translated from German):

Scenario 1:

"Imagine you are at a train station to board a train. A ticket costs 20€. Suddenly, you realize that you have forgotten your wallet at home. Hence, you do not have any money on you to buy a ticket. As you attempt to return home to pick up the wallet, the arrival of your train is announced. You are already running late and the next train is scheduled for 30 minutes later, so you would certainly be late for an important meeting. Instead of picking up the money, you could also try to free-ride on the train. *[LOW DETECTION PROBABILITY: There is no conductor nearby, and until now you have never had to show your ticket.] [HIGH DETECTION PROBABILITY: When entering the train station, you had walked past the conductor.]*"

Scenario 4:

"Imagine someone broke into your home and stole your belongings. A few days before the housebreaking, you had broken your fairly new laptop, which was worth about *[LOW FINANCIAL BENEFIT: 750€] [HIGH FINANCIAL BENEFIT: 2500€]*. According to an expert in an electronics retailer, the laptop cannot be fixed. Consequently, the laptop was worthless when it was stolen. However, since your laptop was sitting in its case on your desk, the burglars were unable to see that it was broken and stole it. You are filling in the damage report to your insurance, on which you have to list the stolen goods and their current value. According to your insurance policy, you have to list the actual value of the goods at the time they were stolen."

Translations of the other four scenarios can be found in Appendix A of the main paper. To assess convergent validity of the scenarios, we additionally applied established measures for delinquency. Specifically, we applied the six-item delinquency scale adapted from Tyler (1990) (1 = never to 4 = often; Appendix B), and a Rasch-scaled delinquency scale by Raudenbush et al. (2003), comprised of 20 binary items of more serious delinquent acts (e.g., have you attacked somebody with a weapon last year? Yes – No; delinquency score range: 0 = no to all, 1 = yes to all). Legitimacy was measured using the six-item *legitimacy of the law* scale (Tyler, 1990) (range: 1 - 4; Appendix B). Convergent validity was assessed using measures for concepts that should be related to legitimacy. More precisely, we measured the construct Attitudes Toward the Criminal Legal System (ATCLS) using an adaptation of a pertinent scale (Martin & Cohn, 2004). The ATCLS scale consisted of 21 items (three items from the original scale were dropped due to incompatibility with the German legal system). Responses were given on a 5-point scale where low values indicate little agreement with the items. Furthermore, we assessed subjective rule of law in Germany based on a definition by Carothers (1998) as a proxy for procedural justice using a self-developed six-item scale. Responses were scored on a 7-point scale with low values indicting little agreement and,

hence, low procedural justice. The items of both scales are listed in Appendix B of the main paper. [1]

Procedure. The experiment was run using an online survey-platform. Throughout the experiment, participants could not go back to change responses on previous pages. Scenarios were presented in random order and participants responded to all scenario-related questions immediately after reading the respective scenario. First, participants were instructed to imagine themselves as being in the situation of the scenarios. Our main dependent variable *probability of committing the crime* was measured using the neutrally phrased question: *How likely are you to [show the critical behavior; e.g., enter the train without a ticket]?* Answers were given on a scale from 0 (not likely) to 100 (very likely) using a scroll bar, which we transformed into a probability score ranging from 0 to 1 for the analysis. On the next page, participants indicated a binary decision concerning the critical behavior: *Honestly, are you going to [show the critical behavior; e.g., enter the train without a ticket]?*

On the following pages, participants indicated their expected detection probabilities and fines for the criminal act. First, participants were asked: "Please imagine that 1000 individuals in Germany *[show the critical behavior; e.g., enter the train without a ticket]*. How many of them get caught?" From the indicated frequency, we calculated the subjective detection probability, which was used in the analyses. Second, we measured the *expected severity of punishment* in case of detection. Participants indicated whether they expected a prison sentence or a fine (in all cases fines should be expected according to German law). Then, participants specified the punishment in terms of duration of incarceration (in days) or fine (in Euros, *expected fine*).[2]

[1] All scales were translated into German by a German native speaker and then translated back into English by an English native speaker. Discrepancies between the original and the retranslated version were discussed and ruled out.

[2] Participants were also asked about the perceived negative utility (in monetary terms) caused by incarceration by indicating how much money they would be willing to pay per day to avoid imprisonment. If a participant expected a prison sentence, we were able to calculate the negative monetary value of a prison sentence based on the number of days they expected to go to prison and the money they would be willing to pay per day to avoid imprisonment. Prison sentences were selected in four cases only (1 x free-riding; 1 x illegal downloading; 2 x

Following the scenarios, participants completed the Tyler delinquency scale, the Raudenbusch delinquency scale, the Tyler legitimacy scale, the self-constructed rule of law scale, the ATCLS scale and indicated demographics.[3]

Results

Descriptive statistics and intercorrelations of scales as well as their (zero-order) correlations with the overall tendency to break the law are presented in Table 1. Participants indicated an average probability of breaking the law of .47 with a wide variation of scores between participants (.17 to .79); nobody indicated being certain to obey the law in all of these scenarios. On the delinquency scale by Tyler, which measured less severe acts of breaking the law, subjects indicated an average score somewhat below 2, which implies that participants performed the acts described in the scale only rarely (less than 5 times). In the Raudenbusch delinquency scale, which also includes more severe crimes, people on average indicated to have been involved in 2 out of 20 instances of criminal behavior.

The reliability of the tendency to break the law across scenarios was low (see main diagonal in Table 1).[4] Low reliability was found for the other two delinquency scales as well. These results indicate that the participants' tendency to break the law is partly content-specific (e.g., some participants would not litter, while they would indeed free-ride). Convergent validity concerning our dependent measure, in terms of correlations between the probability of breaking the law score across our scenarios and the self-reported delinquency measures, was also low, but all correlations were in the expected directions. Correlations between the continuous probabilistic measure and the binary yes/no-measure for breaking the law were

insurance fraud). These cases are not included in the reported analyses. Including the calculated monetary values for these cases did not change the general pattern of results.

[3] The questionnaire included additional control measures that did not reveal further insights, but are available upon request.

[4] Note, however, that this measure might somewhat underestimate reliability, since it does not correct for the variance due to the between-subject manipulations of scenarios.

very high (r = .85), as we had expected. All analyses are reported for the continuous probability measure only.

The average score on the legitimacy scale (range: 1 - 4) indicated that participants largely perceived the legal system to be legitimate; however, considerable variation was observed. The legitimacy scale showed medium reliability and correlated moderately with the related measures ATCLS and rule of law, providing some support for the construct validity of the scale (Table 1).

Table 1. Overall descriptive statistics and bivariate correlations of scales in the Pilot Study.

	M	SD	Min	Max	1	2	3	4	5	6
1 *p*(breaking the law)	0.47	0.15	0.17	0.79	.34					
2 delinquency (Tyler)	1.91	0.42	1.00	3.17	.19	.51				
3 delinquency (Raud.)	0.10	0.09	0.00	0.36	.23	**.54**	.31			
4 Legitimacy	2.80	0.45	2.00	3.67	-.18	**-.41**	**-.29**	.60		
5 ATCLS	2.72	0.41	1.81	3.71	-.15	-.24	-.21	.27	.83	
6 rule of law	5.34	0.82	3.00	7.00	-.02	-.27	**-.35**	**.34**	**.50**	.80

Note. Significant correlations at $p < .05$ are in bold; scale reliabilities (Cronbach's α) are presented in the diagonal of the matrix; *p*(breaking the law) refers to the average probability to break the law per person across all scenarios; $N = 51$.

Table 2 provides descriptive results for the participants' tendency to break the law split by scenario and by our between-subjects manipulation of scenarios. The tendency to break the law varied substantially between scenarios, with only very few subjects littering, but most participants committing insurance fraud. Estimated detection probabilities were highest for parking and free-riding on the train, but lowest for littering, illegal download, and insurance fraud. Similarly, participants' expected fines varied within reasonable ranges. Interestingly, the expected values of fines (=detection probability multiplied by expected fine) were low for all scenarios, indicating that the fine on its own should not be deterring in most cases.

Table 2. Descriptive statistics of probability to break the law, perceived detection probability, expected fine, and expected value of the fine by scenario in the Pilot Study.

	p (breakLaw)		detection probability			exp. fine (in Euro)			expected value of fine		
scenario	*M*	*SE*	*Md*	*M*	*SE*	*Md*	*M*	*SE*	*Md*	*M*	*SE*
1 train lp	.65	.08	.15	.25	.05	40	46.09	2.33	6	11.72	2.78
hp	.49	.07	.20	.23	.04	40	45	2.38	9	10.3	1.66
2 download	.65	.05	.01	.09	.02	400	1168	294.4	10	117.7	62.4
3 speeding	.33	.04	.10	.17	.03	100	150.3	18.88	10	37.03	11.92
4 insur lB	.66	.06	.01	.12	.05	400	682.8	171.1	5	100.3	85.15
hB	.74	.05	.01	.08	.02	500	787.6	139.2	5	64.25	20.6
5 parking	.42	.05	.20	.33	.04	50	83.31	16.16	10.5	29.22	8.42
6 littering	.12	.03	.00	.06	.02	50	147.2	40.65	0.25	19.37	12.15

Note. lp = low detection probability; hp = high detection probability; lB = low benefit; hB = high benefit; insur = insurance fraud.

We tested our hypotheses using a tobit regression model to account for censoring of probability data. When necessary, we used a cluster correction of standard errors on the participant level to account for repeated measurement.

Our first hypothesis, $H1a_{DetPr}$, stated that the probability of breaking the law should decrease with an increasing detection probability. The hypothesis was tested in a regression for the train scenario in which this factor was manipulated (Table 3, model 1). Specifically, we regressed the probability of breaking the law on a dummy variable coding our manipulation of detection probability (1=high detection probability, 0=low detection probability), while controlling for expected fine and legitimacy. Our subtle manipulation of detection probability significantly reduced the probability of the law being broken, thus supporting $H1a_{DetPr.}$

$H1b_{Ben}$ predicted that the manipulation of expected financial benefit should increase the attractiveness of breaking the law and hence make delinquent behavior more likely. Remember that our manipulation increased the financial benefit of breaking the law in the insurance fraud scenario by more than factor three. The frequency of the decision to break the

law increased descriptively by 8% for a higher compared to a lower financial benefit (Table 1), but a tobit regression indicated that the financial benefit manipulation did not significantly increase the probability of the law being broken (Table 3, model 2).

Table 3. Regression models predicting probability to break the law in the Pilot Study.

p(break the law)	(1) train	(2) insurance	(3) all scenarios	(4) all scenarios
Legitimacy	-0.198	-0.240*	-0.109	-0.110+
	(-1.03)	(-2.10)	(-1.61)	(-1.65)
detection probability high (1=yes, 0=low)	-0.315* (-2.13)			
benefit high (1=yes, 0=low)		0.0486 (0.46)		
expected fine (in ln(Euro))	-0.0835 (-0.23)	-0.0491 (-1.20)		0.00195 (0.08)
detection probability measured				-0.0241 (-0.15)
Controls			scenario dummies	scenario dummies
Observations	50	48	306	300
pseudo R^2	0.057	0.109	0.233	0.227
Ll	-37.19	-23.02	-188.7	-188.3
ll(null model)	-39.43	-25.84	-246.2	-243.7
subjects / cluster	50	48	51	51

Note. t statistics in parentheses. Results are from tobit regressions. Scenario dummies also include controls for scenario manipulations. Constant omitted.
+ $p < .1$, * $p < .05$, ** $p < .01$, *** $p < .001$

To test $H1c_{PunSe}$ that punishment severity promotes law-abiding behavior, we included an ln-transformed expected fine in the overall regression across all six scenarios (Table 3, model 4). We controlled for scenarios by including scenario dummies (including two separate dummies for the conditions of the manipulated scenarios) and used natural log-transformed fines to reduce the effect of outliers. The expected punishment severity had no significant

effect on the probability of breaking the law; therefore, our data provide no support for the hypothesis.

To test $H2_{Legit}$, which states that the tendency to break the law should decrease with increasing legitimacy, we included legitimacy in the overall regression and found, as expected, that the two were negatively linked in a one-sided test (which we performed because of the directed hypothesis).

Finally, to test $H5a_{IE\ Legit}$, we extended model 4 in Table 3 by additionally including the two two-way interactions between expected fine and detection probability, on the one hand, and legitimacy, on the other. None of the effects reached conventional significance levels. Only for the interaction with detection probability was there a noticeable tendency that legitimacy reduced and even reversed the negative effect of detection probability (b=.33, $t(51)$=1.03, p=.302).

Discussion

We have explored the influence of the manipulations of detection probability and financial benefit, the predictive value of the expected severity of punishment, and the effect of legitimacy on the probability of the law being broken, all measured using hypothetical scenarios. For most of the scenarios, we have found law obedience rates in the mid-range of the probability scale, precluding floor or ceiling effects. Behavior in the scenarios was related in meaningful ways to two measures of self-reported delinquency, providing convergent validity for our measure, although low internal consistency across scenarios indicated that law obedience might have depended on interactions between specific kinds of crime and persons. Overall, we conclude that the scenarios were appropriate to measure law obedience in a lab setting.

Our results suggest that the probability of the law being broken decreases with an increasing detection probability, as predicted by deterrence theory and as shown in previous studies. Contrary to the prediction of deterrence theory, increases in financial benefit and

expected severity of punishment did not significantly influence the probability to break the law. However, one has to acknowledge that the small sample size resulting in low power impedes the detection of potentially small effects. Finally, as postulated by Tyler (1990), higher legitimacy was related to a lower probability to break the law. The effect was found despite controlling for subjectively expected detection probability and expected fine, indicating that the effects of deterrence factors and legitimacy are partially independent. We did not find evidence for legitimacy and deterrence factors to interact.