Taking shortcuts: Cognitive conflict during motivated rule-breaking

Pfister, Roland and Wirth, Robert and Weller, Lisa and Foerster, Anna and Schwarz, Katharina

University of Würzburg

25 June 2018

Taking shortcuts: Cognitive conflict during motivated rule-breaking

Roland Pfister, Robert Wirth, Lisa Weller, Anna Förster, & Katharina Schwarz

Julius-Maximilians-University of Würzburg

Department of Psychology III

Röntgenring 11

97070 Würzburg, Germany

Accepted Manuscript. Please cite as follows:

Pfister, R., Wirth, R., Weller, L., Foerster, A., & Schwarz, K. A. (2019). Taking shortcuts: Cognitive conflict during motivated rule-breaking. Journal of Economic Psychology, 71, 138-147. doi: 10.1016/j.joep.2018.06.005

Correspondence:

Email: roland.pfister@psychologie.uni-wuerzburg.de

Tel. +49-931-31-81363

Abstract

Deliberate rule violations have typically been addressed from a motivational perspective that asked whether or not agents decide to violate rules based on contextual factors and moral considerations. Here we complement motivational approaches by providing a cognitive perspective on the processes that operate during the act of committing an unsolicited rule violation. Participants were tested in a task that allowed for violating traffic rules by exploiting forbidden shortcuts in a virtual city maze. Results yielded evidence for sustained cognitive conflict that affected performance from right before a violation throughout actually committing the violation. These findings open up a new theoretical perspective on violation behavior that focuses on processes occurring right at the moment a rule violation takes place.

Keywords: rule breaking; optimizing violations; cognitive conflict; executive functions; cheating;

PsycINFO Classification Categories and Codes: 2340 Cognitive Processes; 2360 Motivation & Emotion; 2910 Social Structure & Organization; 3660 Organizational Behavior;

JEL Codes: C91: Design of Experiments: Laboratory, Individual; D81: Criteria for Decision-Making under Risk and Uncertainty; D83: Search; Learning; Information and Knowledge; Communication; Belief; Unawareness; D91: Role and Effects of Psychological, Emotional, Social, and Cognitive Factors on Decision Making;

Data availability: The raw data and corresponding analysis scripts are openly available on the Open Science Framework, osf.io/v8usj

Human agents are motivated to minimize the energy they have to invest in performing a task at hand, and reaching this goal sometimes implies that agents come up with solutions that do not necessarily comply with accepted protocols, norms, and rules.

Such deliberate rule violations have been recognized as a prevalent issue in the context of workplace- and safety-related behavior, and previous studies aimed at uncovering the organizational, personal, and situational factors that give rise to rule-violations (Berry, Ones, & Sackett, 2007; Jacobsen, Fosgaard, & Pascual-Ezama, 2018; Reason, 1990; Yap, Wazlawek, Lucas, Cuddy, & Carney, 2013). A prominent class of violation-producing conditions includes moral considerations like moral licensing, misperceptions of possible hazards and inattention to moral standards, thus highlighting psychological processes that may bias decision-making toward rule violations (Mazar, Amir, & Ariely, 2008; Moore & Gino, 2015; Reason, 1995).

Research that aims at predicting rule violations as a function of certain violation-producing conditions focuses on the binary outcome of observing whether or not a given agent violates a given rule. This approach has been highly successful in applied studies on rule-violation behavior because it can inform decision-making in the field (Mazar & Ariely, 2006; Parker, Reason, Manstead, & Stradling, 1995; Runciman, Merry, & Walton, 2007). At the same time, however, this approach does not allow for a precise and comprehensive understanding of rule violations from a psychological perspective because it does not address the cognitive, motivational, and affective processes that are at work for an individual agent right at the moment that they violate a rule. Recent studies have therefore begun to explore an *agent-centered approach* on deliberate rule violations (Pfister, Wirth, Schwarz, Steinhauser, & Kunde, 2016a; Wirth, Pfister, Foerster, Huestegge, & Kunde, 2016).[1] These studies have documented sustained cognitive conflict during rule-violation behavior that arises due to a continued representation of the rule. Conflict became evident in analyses of movement trajectories that were attracted to the rule-based response option in case of rule violations, and also in electrophysiological

[1] We have previously labeled the two approaches to rule violations as adopting either a third-person or a first-person perspective, with "third-person" referring to studies that assess predictors and precursors of rule-violation behavior as well as its observable consequences, and "first-person" referring to the study of psychological processes during the act of committing a rule violation (Jusyte et al., 2017; Wirth, Foerster, Rendel, et al., 2018). The label of an agent-centered approach for the latter type of studies is used here following suggestions that were raised in the review process.

measures that suggested less direct response retrieval for rule violations than for rule-based responses (Pfister et al., 2016b). Measures of cognitive conflict were further correlated with the likelihood of deciding for rule violations across participants, with larger costs going along with fewer rule violations.

Previous studies on cognitive conflict during deliberate rule violations focused on the violation of simple classification rules (Jusyte et al., 2017; Pfister et al., 2016a; Wirth et al., 2016). This focus allowed for studying the minimal defining feature of rule violations; that is: knowing the behavior that is prescribed by a rule but deliberately performing an alternative course of action.[2] Participants in these studies classified targets based on an arbitrary mapping rule by moving the mouse cursor from a home area in the bottom center of the computer screen to a target area in the upper-left or upper-right corner of the screen. The mapping rule indicating the correct response to each stimulus was instructed at the beginning of the experiment, but participants were encouraged to break this rule from time to time during the experiment by deliberately performing an incorrect movement. Such a setup provides a principled approach to cognitive processing during rule-violation behavior, but at the same time this design choice comes with the limitation of omitting motivational contributions to rule-breaking (with rule violation being embedded in the "meta-rule" of breaking the existing mapping rule at times; Gozli, 2017).

Experimental approaches that aim at isolating elementary process components such as cognitive conflict come with a lasting tradition in psychology, though recent work has called for a more holistic approach to the phenomena under investigation (Gozli & Deng, in press; Kingstone, Smilek & Eastwood, 2008). The present study followed the latter spirit and aimed at investigating cognitive conflict during unsolicited, motivated rule violations, thus providing a bridge between basic, cognitive approaches and applied and economic approaches (cf. van Kleef, Wanders, Stamkou, & Homan, 2015; Verschuere & Shalvi, 2014).

[2] A similar argument can be made for studies that aim at isolating specific cognitive processes involved in lying (e.g., Debey, De Houwer, & Verschuere, 2014; Spence et al., 2001). We will get back to the topic of lying in the General Discussion.

Rule violations can be motivated by a broad range of factors. Economic studies of rule-violation behavior have typically focused on cheating by investigating situations in which individuals can violate a rule or norm in order to attain monetary advantages (Dai, Galeotti, & Villeval, 2018; Fischbacher & Föllmi-Heusi, 2013; Gächter & Schulz, 2016; Gneezy, 2005; Hilbig & Hessler, 2013). Participants are thus motivated to either increase their payoffs or to prevent monetary losses in these situations (Schindler & Pfattheicher, 2017). Studies in workplace- and safety-related settings, by contrast, have often focused on non-monetary motives by investigating shortcutting behavior that is typically labelled as a routine or optimizing violation (Dommes, Granié, Cloutier, Coquelet, & Huguenin-Richard, 2015; Kimbrough & Vostroknutov, 2015; Reason, 1990; Runciman et al., 2007). Routine and optimizing violations both describe behavior in which agents depart from an operating procedure or rule to render the task more enjoyable. Routine violations mainly comprise situations in which the agent short-cuts one or more steps that would be required by a protocol in order to expedite task performance, whereas optimizing violations typically describe situations in which the agent performs unusual actions to enrich a low-demand task (“violations for kicks”; Reason, 1995; Runciman et al., 2007). We will use a broader connotation of the term *optimizing violations* in the following to refer to both situations.

As a first step towards investigating cognitive conflict for motivated rule violations, we opted to study optimizing violations in an applied setting: taking forbidden shortcuts while navigating in traffic. An advantage of studying such rule violations is that traffic rules are explicitly defined which renders forbidden shortcutting a salient event. To measure this type of rule-related shortcutting behavior, we asked our participants to take control of a virtual bicycle courier delivering a pizza in a two-dimensional city map as shown in Figure 1. The only instruction was to deliver the pizza as quickly as possible and participants were informed that they could leave as soon as the last pizza had been delivered. Crucially, we implemented one-way roads in some of the maps, and violating these one-way roads could speed up the delivery at times. Accordingly, we expected participants to be motivated to use these shortcuts (i.e., to perform optimizing violations) and studied whether they would experience cognitive conflict in these situations. We further expected participants to differ substantially regarding their frequency of rule-violations, following findings on rule and norm violations in terms of cheating and lying (DePaulo, Kashy, Kirkendol, Wyer, & Epstein, 1996, Halevy, Shalvi, & Verschuere, 2014; Kimbrough & Vostroknutov, 2015; Mazar & Ariely, 2006), and this frequency difference should be related to cognitive conflict, with strong cognitive conflict going along with fewer decisions in favor of violating rules.

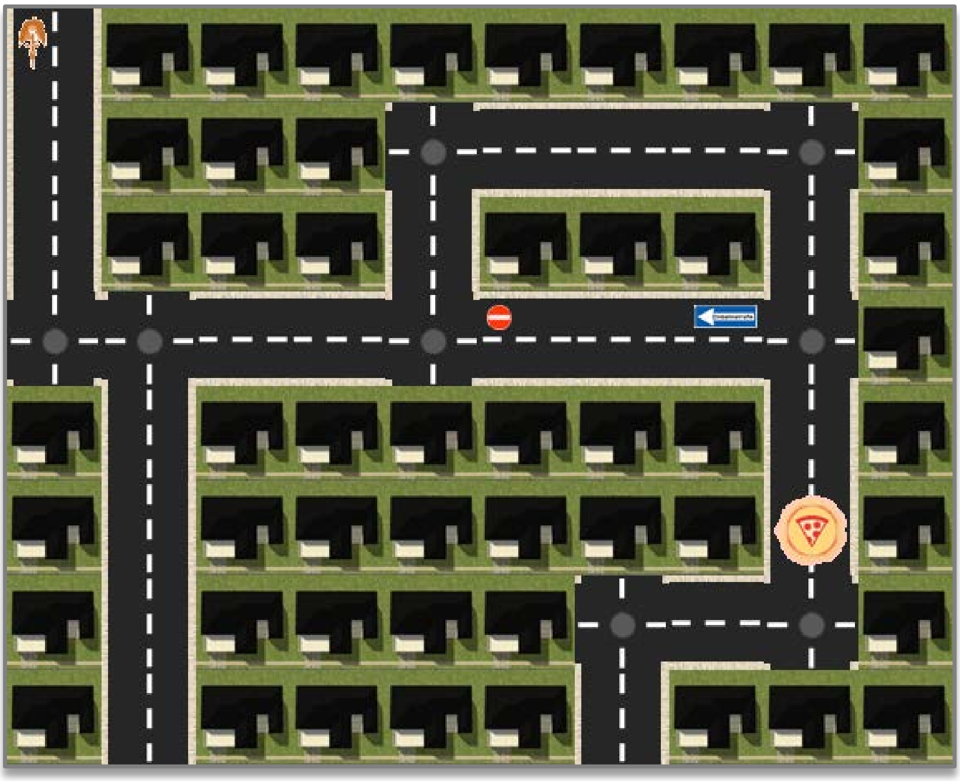

Fig. 1. Exemplar trial of the pizza task to measure violation behavior. Participants navigated a bicycle courier (here: top-left corner) to deliver a virtual pizza (goal location; bottom-right corner) and each keypress moved the courier on the road for one tile (10 x 8 tiles in total). Some roads could be designated as one-way and shortcutting these roads in the forbidden direction could speed up task performance.

Experiment 1

The main question of Experiment 1 was whether cognitive conflict during rule violation would emerge for unsolicited, motivated rule violations such as taking a forbidden shortcut (Hypothesis 1). Cognitive conflict can be assessed in the current experimental paradigm by analyzing the participants' inter-keystroke intervals while performing the tasks: Entering a one-way road in the forbidden direction should induce conflict which should temporarily slow down responding, indicated by prolonged inter-keystroke intervals (see Logan & Crump, 2010, for a similar method applied to typing behavior). As a second question, we assessed whether the strength of this conflict would be related to the individuals' tendency to violate rules (Hypothesis 2). Following previous findings on cognitive conflict during deliberate rule violations (Pfister et al., 2016a), we expected stronger conflict to go along with fewer rule violations as assessed by a correlation with the frequency of rule-violation behavior across participants.[3]

Method

Participants and power analyses

Seventy-two undergraduate psychology students participated for course credit (61 females, 8 left-handers). Their mean age was 20.5 years (range: 18-29 years). This sample size ensured a high power of $1\text{-}\beta > .99$ for the effect size reported in previous studies (e.g., $d_z = 0.95$ for the effect of rule compliance on initiation times in the "violation group" of Exp. 1 in Pfister et al., 2016a). Assuming that the less controlled setting of the present experiments reduces the effect size to a medium effect of $d_z \geq .50$, this sample would still imply a power of $1\text{-}\beta \geq .99$ for detecting relevant conflict effects. Finally, the chosen sample sizes allowed for a power of $1\text{-}\beta = .80$ for detecting correlations of at least medium size ($r \geq .30$). Power calculations were done using the native "power.t.test" and the "pwr.r.test" function of the "pwr" package version 1.1-3 running in R3.3.0. For all power analyses, we assumed $\alpha = .05$ and a directional test of our main hypotheses (note that we

[3] As an additional research question, we explored whether the individual proneness to commit rule violations in the pizza task could be predicted by the individual's tendency to employ cognitive shortcuts as measured in an unrelated categorization task (Pashler & Bayliss, 1991). This was not the case. A more detailed theoretical justification and description of these analyses can be found in the Supplementary Material.

still report two-tailed rather than one-tailed tests to follow common reporting standards). One participant partly guessed the purpose of the experiment and was replaced. The study protocol was approved by the local ethics committee.

Pizza task: Measuring optimizing violations

For the pizza task, participants responded with the four arrow keys of a standard German QWERTZ keyboard to navigate a bicycle courier through city-like 2D-mazes (see Fig. 1). Mazes consisted of 10 x 8 tiles (1.5 cm x 1.5 cm) and each map contained roads, non-passable houses, and a goal location that was signaled by a pizza icon. Some maps additionally contained one or more designated one-way roads.

Pressing a key moved the courier forward one tile and the bicycle movement was always coded relative to the (global) map rather than the (local) courier orientation, i.e., pressing the left arrow moved the bicycle one tile to the left on the screen, irrespective of the bicycle's orientation. The program logged responses, inter-keystroke intervals, and corresponding bicycle locations throughout the trial. The trial ended as soon as the bicycle reached the goal location. The final map stayed on screen for 500 ms and the next trial started after an additional interval of 1000 ms.

The experiment started with a training block of five maps that did not contain any one-way roads and participants were not informed about these upcoming stimuli. Then, the experimenter left the room and the participant worked through two blocks of 60 trials each. The two blocks used the same maps in a fixed sequence. Overall, the participants thus completed 120 trials, 30 of which did not contain any one-way roads, 20 contained one-way roads that did not help to cut short to the goal location if used in the forbidden direction, and 70 contained one-way roads that helped to cut short to the goal location by violating the indicated direction.

Participants further performed a short additional task to measure their tendency to employ cognitive shortcuts in a categorization task (Pashler & Baylis, 1991) and they completed an ad-hoc questionnaire targeting their subjective views on rule-violation behavior after completing the experimental task (see the Supplementary Material for details).

Results

Cognitive conflict during one-way violations

A first analysis aimed at characterizing the distribution of one-way violations across participants (Fig. 2, left panel).[4] Descriptively, this distribution exhibited two separate modes, one at each end of the scale. To quantify this visual impression, we computed two statistics: The bimodality coefficient and Hartigan's dip test (Freeman & Dale, 2013; Pfister, Schwarz, Janczyk, Dale, & Freeman, 2015). The bimodality coefficient amounted to $b = .679$, clearly exceeding the cut-off value of $b_{crit} = .555$ that would be expected for a uniform distribution (Knapp, 2007). Furthermore, the dip test for unimodality (Hartigan & Hartigan, 1985) was significant, $dip = 0.095$, $p < .001$, indicating a non-unimodal distribution.

To test Hypothesis 1, i.e., to probe for cognitive conflict as assessed via inter-keystroke intervals, we distinguished between the very first violation trial and all following violation trials. This was done, because we had chosen not to inform the participants about existence and function of one-way roads during the introduction so that the first act of violation likely involved uncertainty of what to expect when entering the one-way road in the forbidden direction. The analysis could thus only be run for participants who committed at least two violations across the experiment and did not produce any missing data during the first violation, e.g., by reversing direction right after entering the one-way. This procedure resulted in $n = 48$ usable data-sets, thus providing a power of $1\text{-}\beta = .96$ (or $1\text{-}\beta = .92$ when assuming a two-tailed test).

For all participants of the final sample, we calculated mean inter-keystroke intervals for four conditions: (1) keystrokes during a violation trial that were unrelated to the violation itself (i.e., keypresses that were not performed in or right before entering a one-way), (2) keystrokes right before entering a one-way in the forbidden direction, (3) keystrokes that initiated the violation (i.e., entering the one-way), and (4) keystrokes while heading through the one-way in the forbidden direction (see Fig. 2, right panel). Of main

[4] Note that this analysis only included trials in which passing through the one-way road in the forbidden direction served as a shortcut (70 trials per participant). One-way roads for which a violation did not help to shorten the path were used too rarely to allow for meaningful analyses.

interest was the comparison of inter-keystroke intervals when entering a one-way in the forbidden direction as compared to violation-unrelated keystrokes, which provides a direct test for Hypothesis 1. The two remaining keystroke types were mainly included for exploratory analyses. Keystrokes right before entering a one-way allow for evaluating cognitive conflict in a situation in which participants could still turn around and take a rule-conform route. Keystrokes while heading through the one-way, by contrast, allow for assessing behavior while performing a series of consecutive rule-breaking actions (as compared to measures of one instance of rule-violation behavior; e.g., Pfister et al., 2016a).

Inter-keystroke intervals deviating by more than 2.5 standard deviations from their cell mean were considered outliers (3.1%). Because the very first violation of each participant was treated separately, the inter-keystroke interval data were analyzed by a 4 x 2 repeated-measures analysis of variance (ANOVA) with the factors keystroke type (as described above) and violation order (first vs. following violations; see Tab. 1 for complete descriptive statistics).

Most importantly, the described ANOVA revealed a main effect of keystroke type, $F(3, 141) = 12.95$ ($\varepsilon = .51$), $p < .001$, $\eta_p^2 = .22$, driven by slow responses when initiating the violation on the one hand, and short inter-keystroke intervals while passing through the one-way on the other hand (as compared to violation-unrelated responses). Additionally, keystrokes during the first violation trial were overall slower than those of the remaining trials, $F(1, 47) = 33.75$, $p < .001$, $\eta_p^2 = .42$, and the effect of keystroke type was stronger for the first violation than for the remaining violations, $F(3, 141) = 3.23$ ($\varepsilon = .59$), $p = .048$, $\eta_p^2 = .07$. Separate pairwise comparisons indicated that the inter-keystroke interval when entering the one-way was significantly longer than violation-unrelated inter-keystroke intervals for the first violation ($\Delta = 115$ ms), $t(47) = 2.90$, $p = .006$, $d = 0.42$, as well as for the following violations ($\Delta = 40$ ms), $t(47) = 5.94$, $p < .001$, $d = 0.86$. Similarly, inter-keystroke intervals while passing the one-way were significantly shorter than unrelated ones for the first violation ($\Delta = -49$ ms), $t(47) = -3.31$, $p = .002$, $d = -0.48$, and also for the following violations ($\Delta = -25$ ms), $t(47) = -11.10$, $p < .001$, $d = -1.60$. The difference between unrelated inter-keystroke intervals and inter-keystroke intervals right before the violation did not approach significance for either comparison (*ps* > .320).

Tab. 1. Mean inter-keystroke intervals in milliseconds for both experiments and all experimental conditions. Standard errors of paired differences (SE_{PD}) show the within-subject standard error when comparing each condition to the baseline condition of violation-unrelated responses (cf. Pfister & Janczyk, 2013).

		Keystroke Type				SE_{PD}		
	Violation order	Violation-Unrelated	Before Violation	Entering One-Way	Passing One-Way	Before Violation	Entering One-Way	Passing One-Way
Exp. 1	First	255	278	371	209	23	40	14
	Following	202	203	241	176	4	7	2
Exp. 2	First	293	271	349	241	19	22	17
	Following	208	221	263	173	12	15	4

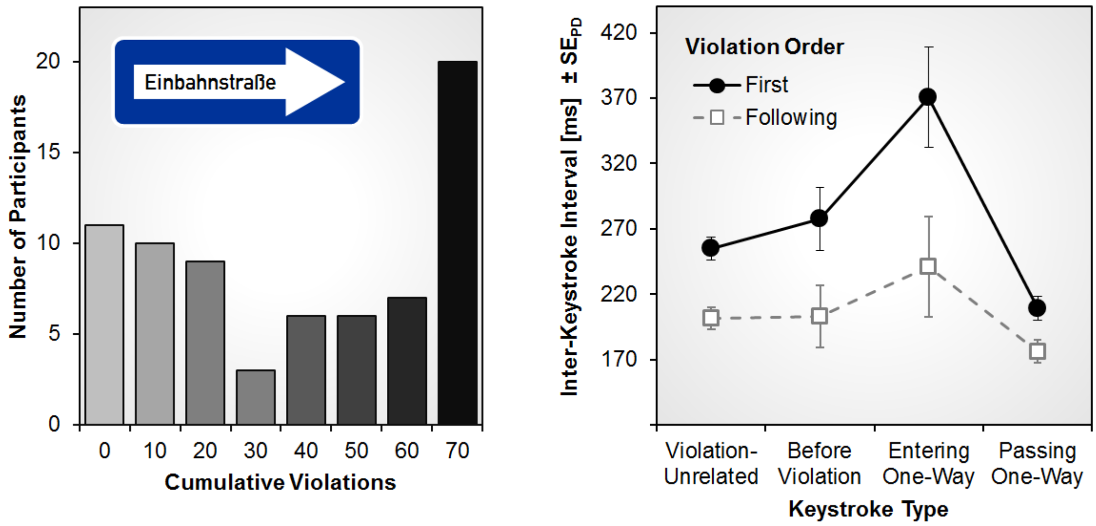

Fig. 2. Central results of the pizza task to measure cognitive conflict during motivated rule-violation behavior in Experiment 1. **Left panel:** Histogram of the individual proneness to violate rules, accompanied by the German road sign for one-way roads. Bins are labelled in terms of their upper boundary and the experimental design permitted up to a maximum of 70 violations. **Right panel:** Inter-keystroke intervals at four different positions during a violation trial of the pizza task. First violation data refer to the very first violation committed in the experiment (when participants did not yet know what to expect when entering the one-way road) whereas the data labeled as "following violations" represents the mean of all subsequent violations. Keystroke types are coded for different responses within a trial in which the participant had committed a violation. Error-bars are standard errors of paired differences (SE_{PD}), computed separately for each keystroke type. For additional data and analyses, see Figure S1 in the Supplementary Material.

Correlation of cognitive conflict and violation frequency

To test Hypothesis 2, i.e., to probe for the hypothesized negative correlation of cognitive conflict during deliberate rule violations and the individual's proneness to violate rules, we computed a conflict index to capture the net effect of rule-violation behavior on performance. To this end we subtracted the mean inter-keystroke interval when passing through a one-way road from the mean interval when entering the road for the repeated violation condition of each participant. To further account for confounds due to variation in overall response speed, we normalized this difference by dividing it by the participant's mean inter-keystroke interval averaged across all four conditions. The conflict index was correlated with the number of one-way violations across participants, $r = -0.29$, $t(46) = 2.05$, $p = .045$ (see Figure S1 in the Supplementary Material). Follow-up tests revealed that this correlation was mainly due to systematically prolonged inter-keystroke intervals when entering the one-way road: When computing separate correlations for the comparison of inter-keystroke intervals when entering the one-way road relative to the baseline of violation-unrelated responses, and for the comparison of inter-keystroke intervals when passing through the one-way, we observed a significant correlation only for the former case (i.e., entering), $r = -0.35$, $t(46) = 2.52$, $p = .015$, but not for the latter case (i.e., passing), $r = 0.03$, $t(46) = 0.20$, $p = .846$.

Discussion

The findings of Experiment 1 lend support to both hypotheses of the present study: Participants showed reliable signs of cognitive conflict when entering a one-way road in the forbidden direction (Hypothesis 1), and the strength of this conflict was negatively correlated with the frequency of rule violation choices across participants (Hypothesis 2). Based on these initial findings, we attempted to replicate and extend the pattern of results in a second experiment.

Experiment 2

Experiment 1 had focused on optimizing violations that are motivated by the desire to expedite task completion. Even though strong effects of cognitive conflict emerged in this setting, it is not clear whether the observation of cognitive conflict would also generalize to other types of motives, especially when rule violations are committed in the face of monetary temptations (Dai et al., 2018; Fischbacher & Föllmi-Heusi, 2013; Gneezy, 2005; Hilbig & Thielmann, 2017). Findings on choice behavior have often suggested monetary incentives to exert a strong motivational pull towards cheating behavior, creating spontaneous impulses toward rule-violation (Bereby-Meyer & Shalvi, 2015; Shalvi, Eldar, & Bereby-Meyer, 2012). These observations could be taken to suggest that the promise of monetary incentives reduces or even overrides cognitive conflict. At the same time, cognitive accounts suggest conflict to arise at a considerably shorter time-scale than motivational processes so that cognitive conflict may also prevail in the presence of monetary temptations (Foerster, Pfister, Schmidts, Dignath, & Kunde, 2013). Experiment 2 therefore replicated the setup of the preceding experiment but introduced monetary incentives – tips for fast deliveries – to study cognitive conflict in such tempting situations.[5]

Hypotheses were as for Experiment 1 and we thus probed for cognitive conflict as measured via inter-keystroke intervals (Hypothesis 1) as well as a negative correlation of the strength of this conflict with the frequency of rule violations across participants (Hypothesis 2).

Method

We performed a direct replication of the pizza task of Experiment 1 with the only addition that participants could earn tips for fast deliveries. Instructions did not mention this added manipulation, and the program decided between fast (tipped) and slow (non-tipped) trials based on an adaptive algorithm. This algorithm ensured that participants were able to receive tips regardless of whether or not they violated rules, though violating improved the chances of obtaining tips in a given trial. To this end, the two experimental blocks were further divided into sub-blocks of 10 trials. For every sub-block, the mean delivery time was computed upon completion, and deliveries were tipped if a delivery was completed faster than the mean minus 1 SD of the previous sub-block. After a tipped delivery, the experiment displayed “You were quick and got a tip” plus their accumulated

[5] We thank the action editor and an anonymous reviewer for stimulating this experiment.

tip that they earned during the whole experiment. To allow for this feedback, the inter-trial-interval was changed from 1000 ms to 3000 ms. Participants earned 1.85 € in tips on average.

Seventy-two new participants were recruited and received either course credit or monetary reimbursement of 5 € (before tips). This sample size ensures a power of $1 - \beta > .99$ for detecting cognitive conflict effects as observed in the inter-keypress intervals of Experiment 1 when assuming similar drop-out as in the preceding experiment. The sample comprised 65 females, 6 left-handers (one participant did not disclose handedness) and the participants' mean age was 26.3 years (range: 19-61 years).

Results

Cognitive conflict during one-way violations

As for Experiment 1, we first examined the distribution of one-way violations across participants (Fig. 3, left panel). This distribution again exhibited two separate modes, one at the lower end of the scale and one at the upper end, though markedly fewer participants opted not to commit a single rule violation. Statistical assessment showed the distribution not to be unimodal as indicated by a bimodality coefficient of $b = .673$, supported by a significant dip-test, $dip = 0.082$, $p < .001$.

To test Hypothesis 1, i.e., to probe for cognitive conflict as captured via inter-keystroke intervals, we again performed a 4 x 2 repeated-measures ANOVA with the factors keystroke type (violation-unrelated, before violation, entering one-way, passing through one-way) and violation order (first vs. following violations; see Fig. 3 and Tab. 1 for corresponding descriptive statistics). A sub-sample of 52 participants was available for this analysis following the same criteria as described for Experiment 1, and we excluded 2.8% of the inter-keystroke intervals as outliers.

Like in Experiment 1, we observed a main effect of keystroke type, $F(3, 153) = 15.67$ ($\varepsilon = .60$), $p < .001$, $\eta_p^2 = .24$, driven by slow responses when initiating the violation and by short inter-keystroke intervals while passing through the one-way (as compared to violation-unrelated responses). Additionally, keystrokes during the first violation trial were overall slower than those of the remaining trials, $F(1, 51) = 45.70$,

$p < .001$, $\eta_p^2 = .47$, whereas the interaction of keystroke type and violation order did not interact for Experiment 2, $F(3, 153) = 0.98$ ($\varepsilon = .73$), $p = .384$, $\eta_p^2 = .02$. Separate pairwise comparisons indicated that the inter-keystroke interval when entering the one-way was significantly longer than violation-unrelated inter-keystroke intervals for the first violation ($\Delta = 56$ ms), $t(51) = 2.47$, $p = .017$, $d = 0.34$, as well as for the following violations ($\Delta = 56$ ms), $t(51) = 3.78$, $p < .001$, $d = 0.52$. Similarly, inter-keystroke intervals while passing the one-way were significantly shorter than unrelated ones for the first violation ($\Delta = -52$ ms), $t(51) = -3.02$, $p = .004$, $d = -0.42$, and also for the following violations ($\Delta = -43$ ms), $t(51) = -7.79$, $p < .001$, $d = -1.08$. The difference between unrelated inter-keystroke intervals and inter-keystroke intervals right before the violation did not approach significance for either comparison (*p*s > .258).

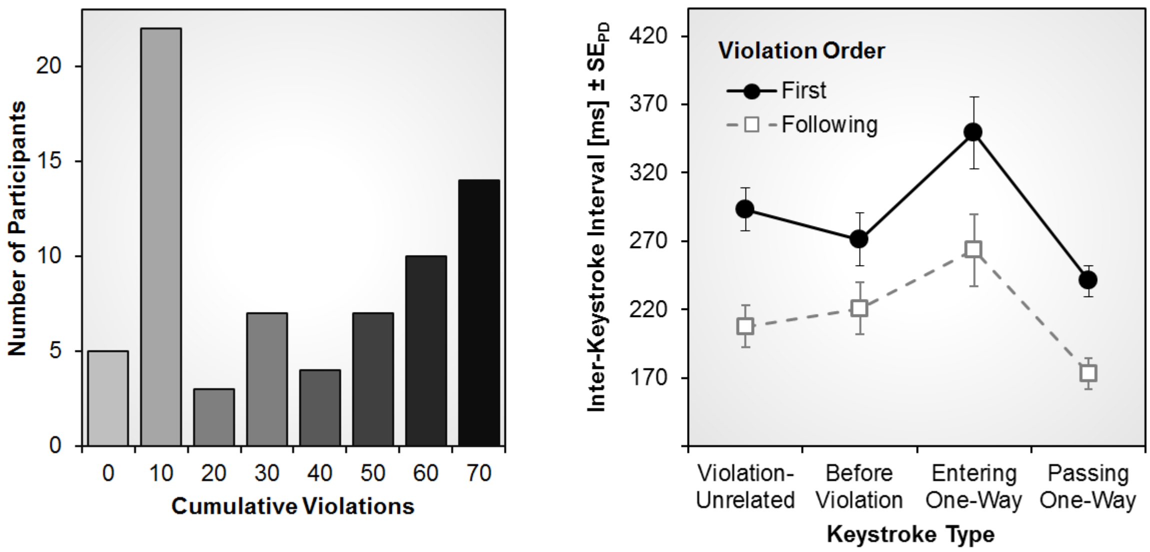

Fig. 3. Results of Experiment 2. **Left panel:** Histogram of the individual proneness to violate rules. Bins are labeled in terms of their upper boundary with a maximum of 70 violations being permitted by the experimental design. **Right panel:** Inter-keystroke intervals for the four different conditions during a violation trial of the pizza task (see also Fig. 2). Error-bars are standard errors of paired differences (SE_{PD}), computed separately for each keystroke type. For additional data and analyses, see Figure S2 in the Supplementary Material.

Correlation of cognitive conflict and violation frequency

To test Hypothesis 2, i.e., to probe for the hypothesized negative correlation of cognitive conflict during deliberate rule violations and the individual's proneness to violate rules, we computed a conflict index as for Experiment 1. The conflict index was again correlated with the number of one-way violations across participants, $r = -0.40$, $t(50) = 3.16$, $p = .003$ (see Figure S2 in the Supplementary Material). Likewise, this correlation mainly derived from prolonged inter-keystroke intervals when entering the one-way road relative to baseline, $r = -0.37$, $t(46) = 2.78$, $p = .007$, and not from the shorter inter-keystroke intervals when passing through the one-way road, $r = 0.15$, $t(46) = 1.08$, $p = .284$.

Discussion

In Experiment 2, participants were able to earn extra money (“tips”) when they delivered a pizza quickly. This additional monetary incentive was introduced to further motivate participants to use forbidden shortcuts by entering a one-way road in the wrong direction on top of the benefit of faster task completion.

Importantly, we still observed cognitive conflict in terms of longer inter-keystroke intervals when participants just entered the one-way road (Hypothesis 1). The results further replicated the negative correlation of the strength of this conflict and the frequency of rule violations (Hypothesis 2), corroborating the results obtained in Experiment 1.

Two observations depart from the previous results, however. First, we did no longer observe an interaction of keystroke type and violation order. For Experiment 1, this interaction had derived from especially large costs when entering a forbidden one-way road for the first time, an effect that is likely due to the uncertainty associated with this response. It seems tempting to attribute this different pattern of results to the stronger motivational pull offered by the additional monetary incentive (Shalvi et al., 2012) which seems to render participants more resilient to uncertainty while it does not overcome the associated cognitive costs.[6] The second observation pertains to the slightly altered shape of the distribution of violation choices across participants: Whereas a sizeable proportion of the participants in Experiment 1 had not committed a single violation throughout the entire session, most participants of Experiment 2 opted to violate the rules at least in a

[6] The first violation condition necessarily comes with rather noisy data especially for the keystrokes before and when entering the one-way, because each participant contributed exactly a single inter-keystroke interval here.

small fraction of the trials. Note, however, that the mean number of violation choices did not differ between experiments as suggested by a post-hoc comparison of both data sets with an average of 33.8 violations per participant in Experiment 1 versus 31.4 violations in Experiment 2, $t(142) = 0.56$, $p = .575$, $d_z = 0.09$. On the one hand, this result may be taken to suggest that the size of the monetary incentives might not have been attractive enough for our participants to instigate violation behavior after receiving tips for fast delivieries. Whether or not decisions for dishonesty depend on the amount of possible payoffs is still under debate at present with several studies showing an impact of payoff magnitude (Gneezy, Rockenbach, & Serra-Garcia, 2013; Hilbig, & Thielmann, 2017) while other studies yielded evidence for the contrary (Fischbacher & Föllmi-Heusi, 2013; Harkrider et al., 2013). The question of whether or not higher payoff magnitudes would alter the cognitive effects of rule-violation behavior thus calls for additional empirical clarification. If one assumes that higher payoffs would not qualitatively alter this pattern, the differing distributions of violation frequencies resonate with the idea that monetary incentives may reduce the impact of uncertainty, thus promoting the chance of observing at least one violation response, while not negating other consequences of rule violation behavior such as cognitive conflict.

General Discussion

The current study set out to bridge cognitive approaches to rule-violation behavior with motivational approaches as they have been put forward in applied psychology and behavioral economics. We studied rule-violation behavior in a task that allowed for short-cutting through one-way roads while participants were to deliver a virtual pizza as quickly as possible. Cognitive conflict during rule-violation behavior was assessed by analyzing the effects of rule violations on continuous task performance while decision biases toward rule-breaking were assessed in terms of the overall frequency of overt rule violations. From a motivational perspective, Experiment 1 focused on optimizing violations, i.e., participants were able to expedite task completion when using one-way roads in the forbidden direction. Experiment 2 built on this setup but introduced additional monetary incentives by offering tips for quick deliveries. We hypothesized that cognitive conflict would emerge also for unsolicited rule violations as operationalized in both experiments (Hypothesis 1) and we further expected the strength of this conflict to go along with fewer instances of rule violation behavior (Hypothesis 2). The results supported both hypotheses and we will discuss these findings in the following.

Conflict and its underlying mechanisms

The analyses of inter-keystroke intervals while participants navigated through the city mazes indicated a systematic slow-down when participants just entered a one-way road in the forbidden direction. We propose that at least for repeated violations, this performance decrement indicates a tug-of-war between automatic tendencies to behave in a rule-based manner (i.e., to turn around and take an accepted route) and the deliberate action plan of moving into the one-way road.

Cognitive research on how rules are represented has indeed indicated that rules are retrieved automatically in the face of rule-related stimuli. This work typically used simple classification rules that prescribed the correct response for certain sets of target stimuli. Encountering any of the stimuli has been shown to retrieve the associated responses even for the very first instance of a stimulus-response episode (Cohen-Kdoshay & Meiran, 2007; Kunde, Kiesel, & Hoffmann, 2003; Wenke, Gaschler, & Nattkemper, 2007), suggesting that rule-based behavior is retrieved automatically even in cases when this behavior does not conform to the agent's current intentions (Dreisbach, 2012). Cognitive conflict during rule-violation behavior thus arises due to the concurrent activation of both, rule-based and rule-violating action tendencies.

The concurrent activation of two opposing action tendencies likely parallels findings on the cognitive psychology of lying, where research has highlighted an initial tendency toward truthful responding that needs to be overcome to successfully tell a lie (Debey et al., 2014; Foerster, Wirth, Kunde, & Pfister, 2017; Spence et al., 2001; for a recent review, see Suchotzki, Verschuere, Van Bockstaele, Ben-Shakhar, & Crombez, 2017). Whether or not this analogy can be taken to suggest similar processing of lying on the one hand and non-deceptive rule violations on the other hand remains to be explored. For instance, motivational accounts have stressed that lying may become the default response given sufficient self-interest in the outcome of the lie (Verschuere & Shalvi, 2014; Bereby-Meyer & Shalvi, 2015). Along the same lines, it has been shown that frequent lying can facilitate dishonest responding to a degree that it appears to become the default response (Van Bockstaele et al., 2012; Verschuere, Spruyt, Meijer, & Otgaar, 2011; for the crucial role of lying recency in this context, see Foerster et al., 2018). Similar results were observed when participants received an explicit false alibi when lying about recently performed actions (Foerster, Wirth, Herbort, Kunde, & Pfister, 2017). By contrast, the violation of arbitrary stimulus-response mapping rules appears not to be malleable to a similar degree and may at times even yield increased cognitive conflict when violations are performed frequently (Wirth, Foerster, Herbort, Kunde, & Pfister, 2018). Similarly, lying typically involves an attempt to conceal the true answer in a communicative setting, which may impose additional processing demands as compared to instances of rule-breaking that do not hinge on communication and successful concealment. Possible commonalities and differences regarding the representation and processing of lying as compared other types of rule- and norm-violation therefore wait for empirical clarification.

Further open questions relate to other potential contributions to the conflict effects observed in the present experiments. It is conceivable that the effects of rule violation on inter-keystroke intervals capture additional factors such as moral considerations relating to the participants' self-image (Mazar et al., 2008; Moore & Gino, 2015). Another process that might contribute to the prolonged inter-keystroke intervals when entering the one-way road is that participants tried to pre-plan the entire movement episode in advance. This assumption might also explain the systematic speed-up when passing through the one-way. Alternatively, or in addition, this speed-up might be attributed to negative affect that has been shown to accompany rule-violation behavior (Wirth, Foerster, Rendel, Kunde, & Pfister, 2018), as participants can be assumed to be motivated to avoid such negative affective states.

Conflict and choice

The distribution of rule violations across participants showed pronounced inter-individual differences with a clear bimodal shape: Participants either used very few forbidden shortcuts or they used many, whereas medium frequencies did not occur as often. This finding is in line with previous individual-differences approaches to cheating, which identified subgroups of mostly honest or “incorruptible” participants that are distinct from other subgroups whose members were more prone to cheating if cheating behavior promised sufficient payoffs (Fischbacher & Föllmi-Heusi, 2013; Hilbig & Thielmann, 2017; Hilbig & Zettler, 2015).

Importantly, the frequency of rule-violation choices was correlated with cognitive conflict as measured via inter-keystroke intervals. This finding resonates with previous observations regarding rule-violation in simple classification tasks (Pfister et al., 2016a). Furthermore, such cognitive conflict has been shown to be absent for convicted criminals, i.e., individuals with a long history of repeated and severe rule-breaking (Jusyte et al., 2017).

In light of these findings, it seems worthwhile to consider the causal mechanisms underlying such correlations. That is: Do frequent violations reduce the associated conflict or, conversely, does anticipated conflict discourage rule-breaking? Regarding the first possibility, frequently committing rule violations has indeed been shown to reduce the cognitive costs associated with this behavior (given that a rule has been violated frequently and just recently), so that this mechanism likely accounts at least for a share of the observed correlation (Verschuere et al., 2011; Foerster et al., 2018). Regarding the second possibility, previous studies have argued that anticipating cognitive conflict may be a driving force behind decisions whether to violate a rule or not (Pfister et al., 2016a). Such an interpretation is tempting also for the present results, especially because it follows recent claims that human agents are highly sensitive to the cognitive effort that has to be invested in an upcoming task (Kool, McGuire, Rosen, & Botvinick, 2010). It thus seems likely that individuals who anticipated stronger cognitive costs are indeed deterred from committing a violation, suggesting that both proposed mechanisms work in concert.

The frequency of rule violations also correlated with the subjective feeling of guilt in the context of rule-violation behavior (at least for Experiment 1; see Supplementary Figure S1 and S2). This finding resonates with theories that propose rule-violation behavior to arise only if the potential gains outweigh negative side-effects related to the agent's self-perception (Hochman, Glöckner, Fiedler, & Ayal, 2016; Mazar et al., 2008; Shalvi, Handgraaf, & De Dreu, 2011). These theories postulate that most human agents intend to maintain a positive and moral self-image. Rule-violation behavior (especially lying and cheating) threatens this self-image and such threats are only condoned if the anticipated gains through a rule violation are sufficiently large. The present study calls for an extension of such psychological frameworks of rule-violation behavior by showing that rule violations do not only entail moral costs but that they also come with robust cognitive costs that emerge right before and while the agent deliberately violates a rule.

Conclusions

The present study shows that unsolicited, motivated rule violations yield cognitive conflict, because agents cannot suppress rule-based tendencies that are automatically activated upon encountering rule-related stimuli. These findings suggest that cognitive conflict is a robust and reliable downstream consequence of rule violation in many different contexts and they promote an agent-centered view on the cognitive, motivational, and affective processes that occur in the acting agent right at the moment a rule violation takes place.

Author note

Experiment 1 was reported in condensed form as part of the first author's PhD thesis (Pfister, 2013), and we are indebted to the dedicated students of the experimental lab course of the winter semester of 2012/13 who performed this experiment. The computer program for the employed pizza task was written during an unexpected overnight stay at Washington Dulles International Airport, due to an apparent lack of usable shortcuts when queueing for customs.

References

Bereby-Meyer, Y., & Shalvi, S. (2015). Deliberate honesty. *Current Opinion in Psychology*, *6*, 195-198.

Berry, C. M., Ones, D. S., & Sackett, P. R. (2007). Interpersonal deviance, organizational deviance, and their common correlates: A review and meta-analysis. *Journal of Applied Psychology*, *92*(2), 410–424.

Blanton, H., & Christie, C. (2003). Deviance regulation: A theory of action and identity. *Review of General Psychology*, *7*(2), 115-149.

Cohen-Kdoshay, O., & Meiran, N. (2007). The representation of instructions in working memory leads to autonomous response activation: Evidence from the first trials in the flanker paradigm. *The Quarterly Journal of Experimental Psychology*, *60*(8), 1140-1154.

Dai, Z., Galeotti, F., & Villeval, M. C. (2018). Cheating in the lab predicts fraud in the field: An experiment in public transportation. *Management Science*, *64*(3), 1081-1100.

Debey, E., De Houwer, J., & Verschuere, B. (2014). Lying relies on the truth. *Cognition*, *132*, 324–334.

DePaulo, B. M., Kashy, D. A., Kirkendol, S. E., Wyer, M. M., & Epstein, J. A. (1996). Lying in everyday life. *Journal of Personality and Social Psychology*, *70*(5), 979-995.

Dommes, A., Granié, M. A., Cloutier, M. S., Coquelet, C., & Huguenin-Richard, F. (2015). Red light violations by adult pedestrians and other safety-related behaviors at signalized crosswalks. *Accident Analysis & Prevention*, *80*, 67–75.

Dreisbach, G. (2012). Mechanisms of cognitive control: The functional role of task rules. *Current Directions in Psychological Science*, *21*(4), 227-231.

Fischbacher, U., & Föllmi-Heusi, F. (2013). Lies in disguise—an experimental study on cheating. *Journal of the European Economic Association*, *11*(3), 525-547.

Foerster, A., Pfister, R., Schmidts, C., Dignath, D., & Kunde, W. (2013). Honesty saves time (and justifications). *Frontiers in Psychology, 4*, 473. doi: 10.3389/fpsyg.2013.00473

Foerster, A. Pfister, R., Schmidts, C., Dignath, D. Wirth, R., & Kunde, W. (2018). Focused cognitive control in dishonesty: evidence for predominantly transient conflict adaptation. *Journal of Experimental Psychology: Human Perception and Performance, 44*(4), 578-602.

Foerster, A., Wirth, R., Herbort, O., Kunde, W., & Pfister, R. (2017). Lying upside-down: Alibis reverse cognitive burdens of dishonesty. *Journal of Experimental Psychology: Applied, 23*(3), 301-319.

Foerster, A., Wirth, R., Kunde, W., & Pfister, R. (2017). The dishonest mind set in sequence. *Psychological Research*, *81*(4), 878-899.

Freeman, J. B., & Dale, R. (2013). Assessing bimodality to detect the presence of a dual cognitive process. *Behavior Research Methods*, *45*(1), 83-97.

Gächter, S., & Schulz, J. F. (2016). Intrinsic honesty and the prevalence of rule violations across societies. *Nature*, *531*, 496-499.

Gneezy, U. (2005). Deception: The role of consequences. *American Economic Review*, *95*(1), 384-394.

Gneezy, U., Rockenbach, B., & Serra-Garcia, M. (2013). Measuring lying aversion. *Journal of Economic Behavior & Organization*, *93*, 293–300.

Gozli, D.G. (2017). Behaviour versus performance: The veiled commitment of experimental psychology. *Theory & Psychology*, *27*(6) 741-758.

Gozli, D. G. & Deng, W. (in press). Building blocks of psychology: On remaking the unkept promises of early schools. *Integrative Psychological and Behavioral Science*. doi: 10.1007/s12124-017-9405-7

Halevy, R., Shalvi, S., & Verschuere, B. (2014). Being honest about dishonesty: Correlating self-reports and actual lying. *Human Communication Research*, *40*(1), 54-72.

Harkrider, L. N., Tamborski, M. A., Wang, X., Brown, R. P., Mumford, M. D., Connelly, S., & Devenport, L. D. (2013). Threats to moral identity: Testing the effects of incentives and consequences of one’s actions on moral cleansing. *Ethics & Behavior*, *23*(2), 133–147.

Hartigan, J. A., & Hartigan, P. M. (1985). The dip test of unimodality. *The Annals of Statistics*, *13*(1), 70-84.

Hilbig, B. E., & Hessler, C. M. (2013). What lies beneath: How the distance between truth and lie drives dishonesty. *Journal of Experimental Social Psychology*, *49*, 263-266.

Hilbig, B. E. & Thielmann, I. (2017). Does everyone have a price? On the role of payoff magnitude for ethical decision making. *Cognition*, *163*, 15-25.

Hilbig, B. E., & Zettler, I. (2015). When the cat's away, some mice will play: A basic trait account of dishonest behavior. *Journal of Research in Personality*, *57*, 72–88.

Hochmann, G., Glöckner, A., Fiedler, S., & Ayal, S. (2015). “I can see it in your eyes”: Biased processing and increased arousal in dishonest responses. *Journal of Behavioral Decision Making*, *29*(2-3), 322-335.

Jacobsen, C., Fosgaard, T. R., & Pascual-Ezama, D. (2018). Why do we lie? A practical guide to the dishonesty literature. *Journal of Economic Surveys*, *32*(2), 357-387.

Jusyte, A., Pfister, R., Mayer, S. V., Schwarz, K. A., Wirth, R., Kunde, W., & Schönenberg, M. (2017). Smooth criminal: Convicted rule-breakers show reduced cognitive conflict during deliberate rule violations. *Psychological Research*, *81*(5), 939-946.

Kimbrough, E. O., & Vostroknutov, A. (2015). Norms make preferences social. *Journal of the European Economic Association*, *14*(3), 608-638.

Kingstone, A., Smilek, D., & Eastwood, J. D. (2008). Cognitive ethology: A new approach for studying human cognition. *British Journal of Psychology*, *99*(3), 317-340.

Knapp, T. R. (2007). Bimodality revisited. *Journal of Modern Applied Statistical Methods*, *6*(1), 8-20.

Kool, W., McGuire, J. T., Rosen, Z. B., & Botvinick, M. M. (2010). Decision making and the avoidance of cognitive demand. *Journal of Experimental Psychology: General*, *139*(4), 665-682.

Kunde, W., Kiesel, A., & Hoffmann, J. (2003). Conscious control over the content of unconscious cognition. *Cognition, 88*(2), 223-242.

Logan, G. D., & Crump, M. J. (2010). Cognitive illusions of authorship reveal hierarchical error detection in skilled typists. *Science*, *330*(6004), 683-686.

Mazar, N., Amir, O., & Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research*, *45*(6), 633–644.

Mazar, N., & Ariely, D. (2006). Dishonesty in everyday life and its policy implications. *Journal of Public Policy & Marketing*, *25*(1), 117–126.

Moore, C., & Gino, F. (2015). Approach, ability, aftermath: A psychological process framework of unethical behavior at work. *Academy of Management Annals*, *9*(1), 235-289.

Mazar, N., Amir, O., & Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research*, *45*(6), 633–644.

Parker, D., Reason, J. T., Manstead, A. S. R., & Stradling, S. G. (1995). Driving error, driving violations and accident involvement. *Ergonomics*, *38*(5), 1036-1048.

Pfister, R., & Janczyk, M. (2013). Confidence intervals for two sample means: Calculation, interpretation, and a few simple rules. *Advances in Cognitive Psychology*, *9*(2), 74-80.

Pfister, R., Schwarz, K. A., Janczyk, M., Dale, R., & Freeman, J. B. (2013). Good things peak in pairs: A note on the bimodality coefficient. *Frontiers in Quantitative Psychology and Measurement*, *4*, 700. doi: 10.3389/fpsyg.2013.00700

Pfister, R., Wirth, R., Schwarz, K., Steinhauser, M., & Kunde, W. (2016a). Burdens of non-conformity: Motor execution reveals cognitive conflict during deliberate rule violations. *Cognition*, *147*, 93-99.

Pfister, R., Wirth, R., Schwarz, K. A., Foerster, A., Steinhauser, M. & Kunde, W. (2016b). The electrophysiological signature of deliberate rule violations. *Psychophysiology*, *53*(12), 1870-1877.

Phipps, D. L., Parker, D., Pals, E. J. M., Meakin, G. H., Nsoedo, C., & Beatty, P. C. W. (2008). Identifying violation-provoking conditions in a healthcare setting. *Ergonomics*, *51*(11), 1625–1642.

Reason, J. (1990). *Human error*. New York, NY: Cambridge University Press.

Reason, J. (1995). Understanding adverse events: Human factors. *Quality and Safety in Health Care*, *4*(2), 80–89.

Runciman, B., Merry, A., & Walton, M. (2007). *Safety and ethics in healthcare: A guide to getting it right*. Aldershot: Ashgate.

Schindler, S., & Pfattheicher, S. (2017). The frame of the game: Loss-framing increases dishonest behavior. *Journal of Experimental Social Psychology*, *69*, 172-177.

Shalvi, S., Eldar, O., & Bereby-Meyer, Y. (2012). Honesty requires time (and lack of justifications). *Psychological Science*, *23*(10), 1264-1270.

Shalvi, S., Handgraaf, M. J. J., & De Dreu, C. K. W. (2011). Ethical manoeuvring: Why people avoid both major and minor lies. *British Journal of Management*, *22*, S16–S27.

Spence, S. A., Farrow, T. F. D., Herford, A. E., Wilkinson, I. D., Zheng, Y., & Woodruff, P. W. R. (2001). Behavioural and functional anatomical correlates of deception in humans. *NeuroReport*, *12*, 2849–2853.

Suchotzki, K., Verschuere, B., Van Bockstaele, B., Ben-Shakhar, G., & Crombez, G. (2017). Lying takes time: A meta-analysis on reaction time measures of deception. *Psychological Bulletin*, *143*(4), 428-453.

Van Bockstaele, B., Verschuere, B., Moens, T., Suchotzki, K., Debey, E., & Spruyt, A. (2012). Learning to lie: Effects of practice on the cognitive cost of lying. *Frontiers in Psychology*, *3*, 526. doi: 10.3389/fpsyg.2012.00526

van Kleef, G. A., Wanders, F., Stamkou, E., & Homan, A. C. (2015). The social dynamics of breaking the rules: Antecedents and consequences of norm-violating behavior. *Current Opinion in Psychology*, *6*, 25–31.

Verschuere, B., & Shalvi, S. (2014). The truth comes naturally! Does it? *Journal of Language and Social Psychology*, *33*(4), 417-423.

Verschuere, B., Spruyt, A., Meijer, E. H., & Otgaar, H. (2011). The ease of lying. *Consciousness and Cognition*, *20*, 908-911.

Wenke, D., Gaschler, R., & Nattkemper, D. (2007). Instruction-induced feature binding. *Psychological Research*, *71*(1), 92-106.

Wirth, R., Foerster, A., Herbort, O., Kunde, W., & Pfister, R. (2018). This is how to be a rule breaker. *Advances in Cognitive Psychology, 14*(1), 21-37.

Wirth, R., Foerster, A., Rendel, H., Kunde, W., & Pfister, R. (2018). Rule-violations sensitize towards negative and authority-related stimuli. *Cognition & Emotion, 32*(3), 480–493.

Wirth, R., Pfister, R., Foerster, A., Huestegge, L., & Kunde, W. (2016). Pushing the rules: Effects and aftereffects of deliberate rule violations. *Psychological Research*, *80*(5), 838-852.

Yap, A. J., Wazlawek, A. S., Lucas, B. J., Cuddy, A. J., & Carney, D. R. (2013). The ergonomics of dishonesty the effect of incidental posture on stealing, cheating, and traffic violations. *Psychological Science*, *24*(11), 2281–2289.

Supplementary Material: Experiment 1

Background

As an additional, exploratory question, we probed whether the individual proneness to commit optimizing violations as measured in the pizza task would not only correlate with cognitive conflict as described in the article, but also whether it may reflect the operation of more basic, low-level processes that have been discussed as cognitive shortcuts. Such shortcuts range from explicit, strategic shortcuts in mental arithmetic (Haider & Frensch, 1996, 1999) and heuristics in decision making (Pachur & Bröder, 2013; Tversky & Kahnemann, 1974) to implicit categorization shortcuts (Pashler & Baylis, 1991).

The different types of cognitive shortcuts have in common that they reflect a consistent tendency to not perform certain operations (deliberate reasoning, rule-based action selection) if the current situations allows for omitting them. Following attempts to ground higher-level processes in the operation of basic perception-action mechanisms such as a potential relation between attentional control and creativity (Zmigrod, Zmigrod, & Hommel, 2015), we tested whether the individual tendency to employ cognitive shortcuts would predict the participant's proneness to rule-related shortcutting behavior. To measure cognitive shortcuts, we assessed implicit categorization shortcuts that are used to avoid effortful response-selection processes (Pashler & Baylis, 1991).

Cognitive shortcuts are typically observed in choice reaction tasks in which participants respond to target stimuli in a succession of trials. In this case, we opted to present the faces of different celebrities as target stimuli and will therefore refer to this task as the *celebrity task*. If the current target stimulus matches the stimulus that was encountered in the preceding trial (stimulus repetitions), responses are typically much faster than with changing stimuli. This finding has been taken to indicate that the time-consuming categorization of the target stimulus is skipped when a stimulus is repeated (Bertelson, 1963; Pashler & Baylis, 1991; see also Tan & Dixon, 2011). Categorization shortcuts can thus be construed as a tendency to not select an appropriate response according to a specific mapping rule, but rather bypass this rule by relying on memory traces that are still active.

The cognitive shortcuts in the categorization task and in the pizza task arguably differ in many regards: Categorization shortcuts take place on a scale of a few hundred milliseconds and they are rarely employed deliberately (Pashler & Baylis, 1991), whereas the type of rule violations studied here takes place on a larger timescale and – assuming that participants are aware of the salient one-way signs – are based on a deliberate decision to violate this rule. Despite these differences, categorization shortcuts and optimizing violations have in common that the agent reaches a desired end – a correct categorization or successful performance, respectively – by other means than suggested by the task at hand. We therefore predicted that the individual tendency to violate rules in the *pizza task* would not only be negatively related to measures of cognitive conflict in this task, but that it would also be positively related to the individual tendency to exploit stimulus repetitions in the *celebrity task*.

Supplementary Method

Celebrity task: categorization shortcuts

For the celebrity task, participants responded with the keys *J*, *K*, and *L* of a standard computer keyboard, operated by the index, middle, and ring finger of the right hand. The keys were marked with colored patches (orange, green, and white) and instructions always referred to these colors. Target stimuli were grayscale portraits of six celebrities (3.5 cm x 3.5 cm) that appeared on a 17'' monitor. All six celebrities were likely to be well-known among German university students: Angela Merkel (German chancellor), Queen Elizabeth

II., Angelina Jolie (actress), Günther Jauch (German quizmaster), Johnny Depp (actor), and Dirk Nowitzki (Würzburgian basketball player). Two portraits, one male and one female, were mapped to each response key and the stimulus-response mapping was counterbalanced across participants.

Each trial simply featured a target stimulus and participants were to respond as quickly and accurately as possible with the assigned key. The stimulus remained on screen until a response was given and wrong responses triggered error feedback for 1500 ms. The next trial started after 500 ms; responses during the inter-trial interval produced an error message.

Participants worked through eleven blocks of 54 trials each (i.e., each stimulus was displayed nine times per block in a random order) and feedback after each block informed the participants about their mean response time (RT) and the number of errors to ensure a high motivation. The first block was considered practice and did not enter the analyses.

Ad-hoc questionnaire

We further administered a short ad-hoc questionnaire (in German language) after the experiment that probed for the participant's attitude towards rule violations. It featured three questions that could be answered on a visual analogue scale (length: 7.1 cm) with verbal anchors at both ends. The first question translated to "If you violate a rule, how guilty do you feel?" ('feeling guilty'; *not very guilty* to *very guilty*) whereas the second question targeted directly how prone participants were to committing violations: "How often do you violate rules?" ('subjective frequency'; *very rarely* to *very frequently*). The final question translates to "How strongly would you condemn others for breaking a rule on purpose?" ('condemn others', *not very much* to *very strongly*).

Supplementary Results

Celebrity task: Cognitive shortcuts

For analysis of the celebrity task, we only considered RTs of correct trials (errors occurred in 5.3% of all trials) and also excluded the first trial of each block and trials that were preceded by errors to avoid confounding effects due to restart costs and error processing. RTs that deviated by more than 2.5 standard deviations from their cell mean were discarded as outliers (3.0%).

The remaining RTs were aggregated to separate means for the three conditions of interest: stimulus repetitions (444 ms), response repetitions (614 ms), and complete alternations (602 ms). These means differed significantly, as indicated by a repeated-measures analysis of variance (ANOVA), $F(2, 142) = 331.84$, $p < .001$, $\eta_p^2 = .82$. The critical effect for the current study, however, was not the omnibus ANOVA but rather the pairwise comparison of complete alternations and stimulus repetitions. Considered separately, this repetition benefit ($RT_{Complete\ Alternation} - RT_{Stimulus\ Repetition}$) amounted to sizeable 158 ms and was significantly different from zero, $t(71) = 22.90$, $p < .001$, $d = 2.70$.

Correlational analyses

As for the evaluation of the participants' proneness to committing optimizing violations and its relation to measures from within the pizza task, we performed correlational analyses between the individual number of violations and four additional predictor variables (see Figure S1 and Table S1).

The first predictor was a repetition index as derived from the celebrity task. We computed this index by normalizing the individual repetition benefits at the participant's mean RT (repetition index = repetition benefit / mean RT * 100).

The remaining three predictors were the ratings for the three questions in the ad-hoc questionnaire (feeling guilty, subjective frequency, and condemn others; measured in % of the visual analogue scale).

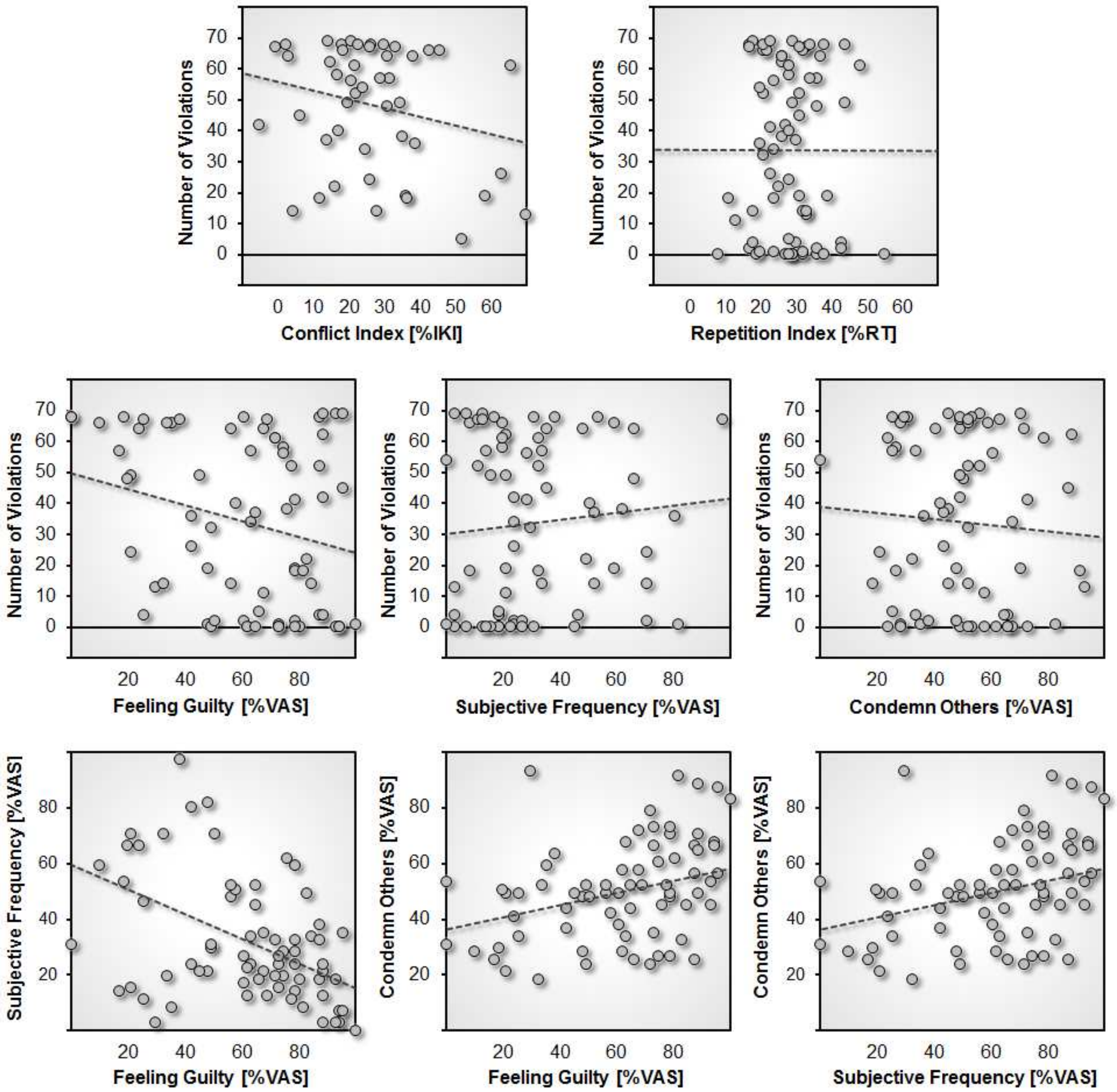

Fig. S1. Results of the correlational analyses. **Upper row:** Bivariate correlations between an individual's proneness to violate rules, the conflict index as a measure of cognitive conflict during rule violations, and the repetition index as a measure for the use of cognitive shortcuts. **Middle row:** Bivariate correlations between an individual's proneness to violate rules and the subjective ratings on the post-experimental questionnaire. **Lower row:** Intercorrelations of the three questionnaire items.

The only significant predictor of the number of violations in this analysis was the subjective guilt when committing violations, r = -0.25, $t(70)$ = -2.16, p = .034, with a regression line equating to $\hat{y}$ = -0.26 • x + 49.65. Accordingly, participants committed less one-way violations, the more they rated themselves to generally feel guilty after having violated a rule. Importantly, the repetition index was not related to the number of one-way violations. To evaluate whether these data can indeed be taken to indicate the absence of a correlation (rather than insensitivity due to the current sample size), we further computed the 95% confidence interval around the two latter coefficients. This was done by applying the

Tab. A1. Pairwise correlations of the repetition index as a measure for the individual proneness to taking cognitive shortcuts, the ad-hoc-questionnaire administered after the experiment, and the number of violations committed by a participant in the pizza task (violation count). The upper diagonal of the table lists correlation coefficients (significant correlations are in italics) whereas the lower diagonal gives the corresponding *p*-values. All correlations are based on the entire sample of *n* = 72 participants of Exp. 1.

	Repetition Index	Feeling Guilty	Subjective Frequency	Condemn Others	Violation Count
Repetition Index		*-0.28*	0.16	-0.01	0.00
Feeling Guilty	*.015*		*-0.49*	*0.30*	*-0.25*
Subjective Frequency	.169	*<.001*		*-0.28*	0.10
Condemn Others	.901	*.010*	*.017*		-0.07
Violation Count	.981	*.034*	.395	.560	

Fisher-Z transformation, computing the confidence interval in Z-space as $Z \pm 1.96 \cdot \frac{1}{\sqrt{72-3}}$, and re-transforming the resulting boundaries to correlation coefficients. This procedure resulted in a 95% confidence interval of [-0.32; 0.14] for the correlation of the repetition benefit (i.e., the raw scores as measured in ms) and the number of violations (r = -0.09) and a confidence interval of [-0.23; 0.23] for the correlation of the repetition index (i.e., the repetition benefit relative to the overall RT level) and the number of violations (r = 0.00). Both confidence intervals clearly exclude the minimum expected value of r = +0.30.

Supplementary Discussion

The tendency to violate rules in the pizza task seems to be independent of cognitive categorization shortcuts as measured in the celebrity task. These observations suggest that deliberate decisions to violate rules might not be traced back to very basic cognitive shortcuts. Thus, at least for the current operationalization of rule violations and cognitive shortcuts, it does not seem as if the deliberate decisions leading to a routine or optimizing violation (Reason, 1990, 1995) drew on rather automatic, low-level shortcuts that bypass certain categorization processes (Pashler & Baylis, 1991; Tan & Dixon, 2011). Instead, the subjective assessment of possible consequences that might result from rule violations (i.e., felt guilt) seems to determine whether an optimizing violation is committed or not.

Aside from methodological problems of interpreting non-significant results, it seems worthwhile to consider a possible alternative explanation that might also explain the present null-correlation between the number of rule violations and the repetition benefit as an index of cognitive shortcuts. Clearly, the number of rule violations is a rather discrete measure of how often a decision process converged on one or the other option (resulting in a rule violation or rule-based behavior). By contrast, the repetition benefit seems to be a continuous, performance-based measure because it is based on differences between mean RTs in two conditions. The null-correlation could thus partly be driven by different information captured by each measure (the outcome of a process vs. the speed of a process). This conclusion seems premature, though. Rather, Pashler and Baylis (1991) argue that the repetition benefit does not indicate a genuine speedup of response selection but rather a shortcut that actually skips response selection processes (for similar views, see Dehaene, 1996; Smith, 1968; Smith, Chase, & Smith, 1973; Tan & Dixon, 2011).

Even though repetition effects are likely to entail additional components (e.g., Soetens, 1998; Sommer, Leuthold, & Soetens 1999), the assumed shortcut would imply a rather discrete mechanism that either takes place (creating a repetition benefit in a given trial) or not. Following this logic, differences in repetition benefits across participants can be seen as a measure of how often a shortcut it used. The applied correlation analysis thus seems to be methodologically sound and the non-significant result might indeed suggest independent processes.

The frequency of rule violations was, however, predicted by the subjective feeling of guilt in the context of rule-violation behavior.[1] This finding is in accordance with theories that propose rule-violation behavior to arise only if the potential gains outweigh negative side-effects related to the agent's self-perception as outlined in the article (Mazar, Amir, & Ariely, 2008).

[1] The correlation coefficient for self-reported guilt and the frequency of committed rule violations was only modest in size. This points to additional influences though the present effect size should also be seen in the context of a possibly limited reliability of single-item data.

Supplementary Material: Experiment 2

Participants of Experiment 2 were not asked to perform the celebrity task so that we restricted the follow-up analyses on correlations between the participants' proneness to violate rules (i.e., their violation frequency) and the three items of the ad-hoc questionnaire.

The questionnaire was identical to the one used in the preceding experiment and contained the questions "If you violate a rule, how guilty do you feel?" ('feeling guilty'; *not very guilty* to *very guilty*), "How often do you violate rules?" ('subjective frequency'; *very rarely* to *very frequently*), and "How strongly would you condemn others for breaking a rule on purpose?" ('condemn others', *not very much* to *very strongly*). Two participants had to be removed from the analysis because they failed to answer the questionnaire so that all correlational analyses of the questionnaire data are based on a sample size of n = 70 participants.

Figure S2 shows bivariate scatterplots for all variable combinations, and Table A2 lists the resulting pairwise correlations. Though the general pattern of results resembled the correlational data of Experiment 1, there were slight differences in terms of which correlation reached the conventional level of significant. That is, in contrast to Experiment 1, subjective guilt only showed a non-significant trend toward a negative correlation with the number of rule violations, whereas moderate correlations emerged for the subjective frequency and condemn others.

To follow up on these somewhat diverging results, we pooled the data of both experiments to re-assess all relevant correlations with higher power. This overall analysis yielded small but significant bivariate correlations of all three questionnaire items with the individual violation frequency; guilt: r = -0.21, $t(170)$ = 2.55, p = .013, subjective frequency: r = 0.27, $t(170)$ = 3.40, p = .001, condemn others: r = -0.19, $t(170)$ = 2.29, p = .024.

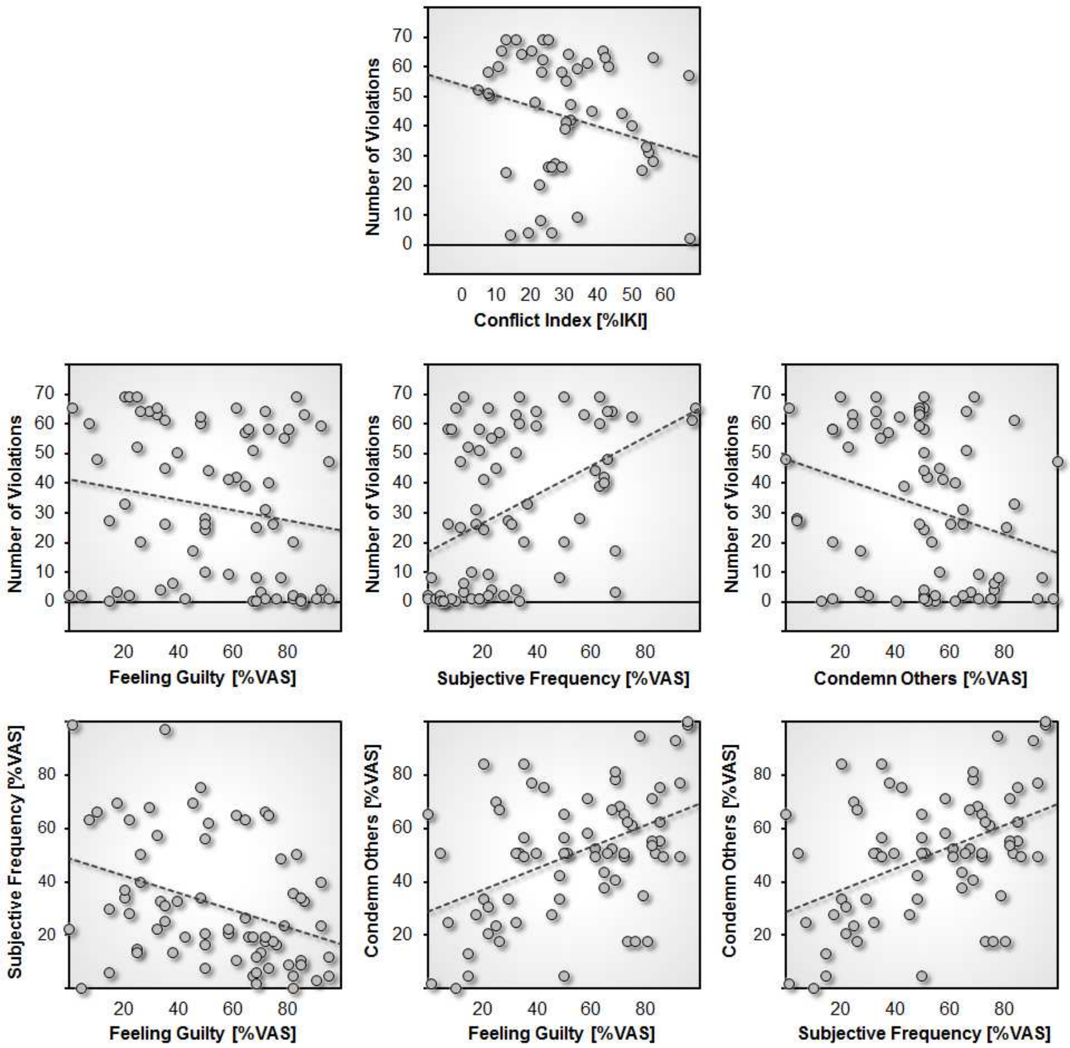

Fig. S2. Results of the correlational analyses of Experiment 2. **Upper row:** Bivariate correlation between an individual's proneness to violate rules and the conflict index as a measure of cognitive conflict during rule violations. **Middle row:** Bivariate correlations between an individual's proneness to violate rules and the subjective ratings on the post-experimental questionnaire. **Lower row:** Intercorrelations of the three questionnaire items.

Tab. A2. Pairwise correlations of the ad-hoc-questionnaire and the number of violations committed by a participant in the pizza task (violation count) of Exp. 2. The upper diagonal of the table lists correlation coefficients whereas the lower diagonal gives the corresponding *p*-values.

	Feeling Guilty	Subjective Frequency	Condemn Others	Violation Count
Feeling Guilty		-0.36	0.45	-0.18
Subjective Frequency	.002		-0.35	0.46
Condemn Others	<.001	.003		-0.29
Violation Count	.134	<.001	.013	

Supplementary References

Bertelson, P. (1963). S-R relationships and reaction time to new versus repeated signals in a serial task. *Journal of Experimental Psychology*, *65*(5), 478-484.

Dehaene, S. (1996). The organization of brain activations in number comparison: Event-related potentials and the additive-factors method. *Journal of Cognitive Neuroscience*, *8*(1), 47-68.

Haider, H., & Frensch, P. A. (1996). The role of information reduction in skill acquisition. *Cognitive Psychology*, *30*(3), 304-337.

Haider, H., & Frensch, P. A. (1999). Eye movement during skill acquisition: More evidence for the information-reduction hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(1), 172-190.

Mazar, N., Amir, O., & Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research*, *45*(6), 633–644.

Pachur, T. & Bröder, A. (2013). Judgment: A cognitive processing perspective. *Wiley Interdisciplinary Reviews: Cognitive Science*, *4*, 665-681.

Pashler, H., & Baylis, G. (1991). Procedural learning: 2. Intertrial repetition effects in speeded-choice tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(1), 33-48.

Reason, J. (1990). *Human error*. New York, NY: Cambridge University Press.

Reason, J. (1995). Understanding adverse events: Human factors. *Quality and Safety in Health Care*, *4*(2), 80–89.

Smith, E. E., Chase, W. G., & Smith, P. G. (1973). Stimulus and response repetition effects in retrieval from short-term memory. *Journal of Experimental Psychology*, *98*(2), 413-422.

Smith, M. C. (1968). The repetition effect and short-term memory. *Journal of Experimental Psychology*, *77*(3), 435-439.

Soetens, E. (1998). Localizing sequential effects in serial choice reaction time with the information reduction procedure. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(2), 547-568.

Sommer, W., Leuthold, H., & Soetens, E. (1999). Covert signs of expectancy in serial reaction time tasks revealed by event-related potentials. *Perception & Psychophysics*, *61*(2), 342-353.

Tan, S., & Dixon, P. (2011). Repetition and the SNARC effect with one- and two-digit numbers. *Canadian Journal of Experimental Psychology*, *65*(2), 84-97.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124-1131.

Zmigrod, S., Zmigrod, L., & Hommel, B. (2015). Zooming into creativity: individual differences in attentional global-local biases are linked to creative thinking. *Frontiers in Psychology*, *6*:1647. doi: 10.3389/fpsyg.2015.01647