Burdens of non-conformity: Motor execution reveals cognitive conflict during deliberate rule violations

Roland Pfistera,⇑, Robert Wirtha, Katharina A. Schwarzb, Marco Steinhauserc, Wilfried Kundea

a Department of Psychology, Julius Maximilians University of Würzburg, Röntgenring 11, 97070 Würzburg, Germany
b Department of Systems Neuroscience, University Medical Center Hamburg-Eppendorf, Martinistr. 52, 20246 Hamburg, Germany
c Catholic University of Eichstätt-Ingolstadt, Ostenstraße 25, 85072 Eichstätt, Germany

Article info
Article history:
Received 12 November 2014
Revised 9 November 2015
Accepted 19 November 2015

Keywords:
Rule violation
Rule representation
Movement trajectories

Abstract
Rule compliance is pivotal for the regulation of social behavior. Still, humans deliberately violate rules at times – be it for personal reasons or for a higher good. Whereas previous research has studied the preconditions and consequences of rule violations, essentially nothing is known about the cognitive processes right at the moment a rule violation takes place. Here we show that merely labeling an action as rule violation induces substantial conflict between rule violation and compliance, as revealed by participants’ bias towards rule-complying motor actions. Moreover, conflict that comes with violating a rule was much stronger than conflict that comes with following an alternative rule, even if both decisions result in the same observable behavior. These observations open a new theoretical perspective on rule violation behavior, shifting the focus toward the cognitive processes operating during the very act of rule violation.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

“I ain’t gonna pay no attention to your rules” sings hard-rock legend AC/DC. Violation of social rules is not confined to hard rock musicians, however. It is a common human phenomenon with sometimes positive consequences, as with moral courage, and sometimes negative consequences as with scientific misconduct (Stroebe, Postmes, & Spears, 2012). Whereas behavioral research has delineated situational and organizational determinants of violation behavior (Phipps et al., 2008; Reason, 1990; Yap, Wazlawek, Lucas, Cuddy, & Carney, 2013), very little is known about the consequences of rule violation right at the moment it takes place. The present experiments are a first step in this direction by showing that even simple motor actions differ depending on whether they aim at following or breaking rules.

Because rules generally trigger compliance (Asch, 1956; Cialdini & Goldstein, 2004; Deutsch & Gerard, 1955; Ruff, Ugazio, & Fehr, 2013; van de Waal, Borgeaud, & Whiten, 2013; Whiten, Horner, & de Waal, 2005) and obedience to authority (Milgram, 1963, 1974), we hypothesized that rule violation inflicts conflict on the rule breaker who is torn between doing what is normally acknowledged and intentionally doing the opposite. Further support for this hypothesis comes from studies on how rules are represented in the human cognitive system and how their representation shapes an agent’s behavior (for an overview, see Bunge & Wallis, 2007). Studies on rule representation often employed controlled stimulus-response (S–R) paradigms to assess how S–R mapping rules are learned and how they are implemented in task sets. A striking result of this line of research is that merely instructing an arbitrary S–R mapping rule will yield automatic response activation upon encountering the associated stimulus (Cohen-Kdoshay & Meiran, 2009; Hommel, 2000; Kunde, Kiesel, & Hoffmann, 2003, 2005; Reisenauer & Dreisbach, 2013; Wenke, Gaschler, & Nattkemper, 2007). Merely instructing a rule thus seems sufficient to automatically retrieve rule-based behavior whenever the agent is in a situation that is relevant to the rule at hand.

While automatic retrieval of rule-based behavior facilitates actions that aim at following the rule, such retrieval obviously hinders any actions that explicitly aim at violating the rule. In this latter case, the agent not only needs to deliberately access the intended action but might also be faced with cognitive conflict (Botvinick, Braver, Barch, Carter, & Cohen, 2001), resulting from the parallel activation of rule-based action plans alongside the intended action (Pfister, 2013).

To reveal such cognitive conflict, we designed a mouse-tracking paradigm in which participants moved a cursor toward a left or right target position according to an S–R mapping rule. Crucially,
they obeyed the mapping rule in some trials but they violated it on occasion. To isolate the effects of labeling a certain behavior as rule violation from further processes that might accompany rule violation behavior, we ensured that the task did not involve any sanctions or otherwise negative consequences. We opted for analyzing movement trajectories of the mouse cursor to assess the impact of the task rule, because such trajectories offer a unique measure of bias towards specific response options (e.g., Freeman & Ambady, 2009, 2011; McKinstry, Dale, & Spivey, 2008; Pfister, Janczyk, Wirth, Dignath, & Kunde, 2014; Song & Nakayama, 2009). Among others, giving deceptive responses to simple yes/no questions was shown to yield a bias toward the honest response option (Duran, Dale, & McNamara, 2010), and, likewise, trajectories were biased toward tempting response alternatives when probing for self-control conflicts (Dignath, Pfister, Eder, Kiesel, & Kunde, 2014). Furthermore, mouse trajectories have been found to be sensitive to internal representations such as anticipated action consequences (Pfister et al., 2014; Wirth, Pfister, Janczyk, & Kunde, 2015). Spatial characteristics of the performed mouse movements thus appear as a prime measure to assess a possible impact of rule representations during the act of rule violation, and we hypothesized the corresponding trajectories to be attracted toward the rule-based response option. Two experiments provided compelling evidence for a profound impact, both, when participants were able to decide whether to violate or not (Experiment 1), and when violations were prompted externally (Experiment 2).

2. Experiment 1: Choosing to violate

We compared two groups of participants: a violation group and a reversed rule group. Participants of the violation group were instructed with one specific S–R mapping rule but were asked to indicate whether or not they wanted to follow the rule before each trial. Participants in the reversed rule group, by contrast, received slightly changed instructions: They were presented with two tasks with oppositional mapping rules and they indicated whether they would perform the original task or the task with opposite mapping. Thus, one and the same motor action was labeled as rule violation group by framing the study as investigating task performance when working on two different tasks. The instruction screen therefore presented both mappings simultaneously with the labels “Task 1” and “Task 2”. Also in contrast to the violation group, participants of the reversed rule group were asked to choose whether to perform Task 1 or Task 2 at the beginning of each trial and to choose spontaneously between these two options. All other aspects of the task were as described for the violation group.

2.1. Methods

2.1.1. Participants, apparatus, and stimuli

We recruited 20 participants for the violation group (mean age = 20.5 years, 14 female, 2 left-handed) and another 20 participants for the reversed rule group (mean age = 21.4 years, 17 female, 2 left-handed). Handedness was determined by self-report and participants operated a standard computer mouse with their right hand and placed their left hand on the arrow keys of the keyboard. Stimuli appeared on a 17” computer monitor at a viewing distance of about 60 cm. Target stimuli were two astrological symbols (Aries vs. Gemini, displayed in 60 pt. MS Gothic font), mapped to a left and a right response, respectively. The S–R mapping was counterbalanced across participants. That is: For one half of the participants, the Aries symbol prompted a movement to the left target area whereas the Gemini symbol prompted a movement to right target area; for the other half of the participants, the Gemini symbol prompted a movement to the left target area whereas the Aries symbol prompted a movement to right target area.

2.1.2. Framing and instructions

At the beginning of the session, participants were introduced to the concept of home and target areas and were given time to acquaint themselves with the setup by moving the cursor between the home area and the target areas. When they felt confident to continue, the experimenter pressed the space bar, which made the two possible target stimuli appear simultaneously on the screen. The experimenter then told the participant that the following task would revolve around a rule that mapped the two target stimuli to a left or right response. Participants could terminate this display by moving their cursor to one of the target areas which cleared the screen, followed by the task rule which was displayed in large (40 pt.) font.

During this initial framing, participants in the violation group were informed that the study investigated the impact of rules on behavior and that they were to work according to a single task rule. The experimenter took care to stress the word “rule” and to avoid alternative terms to describe the task such as “stimulus classification” or “categorization”. After memorizing the mapping rule, participants were informed that they could choose before every trial whether they wanted to follow the rule or whether they wanted to violate the rule and intentionally commit an error. The experimenter also asked the participants to decide spontaneously between rule following and rule violation without using any specific strategy. To conclude the instructions, a summary screen showed four bullet points that described the experimental procedure, again emphasizing that participants would indicate to either follow the rule or violate it and commit an error by intention.

The instructions of the reversed rule group differed from those of the violation group by framing the study as investigating task performance when working on two different tasks. The instruction screen therefore presented both mappings simultaneously with the labels “Task 1” and “Task 2”. Also in contrast to the violation group, participants of the reversed rule group were asked to choose whether to perform Task 1 or Task 2 at the beginning of each trial and to choose spontaneously between these two options. All other aspects of the task were as described for the violation group.

2.1.3. Procedure

Participants completed a training block and 8 experimental blocks of 50 trials each, 25 trials with Aries and 25 trials with Gemini as target stimulus (see Fig. 1 for a schematic of the trial procedure). At the beginning of each trial, participants indicated their current intention (“compliance response”). For the violation group, the corresponding display featured the correct task mapping in the upper half of the screen. The lower half of the screen showed two boxes containing the German words “Korrekt” (Eng. correct) and “Fehler” (Eng. error), to indicate both options that were available to the participant to choose from. We chose the label “error” instead of “violation” to further stress that this behavior was not in accordance with the still active rule and to avoid misunderstandings in terms of applying the reversed rule. The locations of the correct option and the error option were counterbalanced across participants but constant across trials for each individual. Participants responded whether they would comply to the mapping rule by pressing either the up-key or the down-key on the computer keyboard with their left hand.
point onward, we sampled the
features of the composite action and the rule. Thus, choosing Task 2 was identical to choosing to violate the mapping rule (i.e., committing an error by intention) except for the different framing of the response.

Participants gave the compliance response at leisure – correct vs. violation or Task 1 vs. Task 2, respectively. Then, the screen was blanked and three areas appeared: The home area in the bottom center and the two target areas to the upper left and upper right of the screen. From this point onward, the mouse cursor was displayed as a small circle (0.5 cm in diameter) and the program waited for the participants to move inside the home area. Each area measured 1.6 cm in diameter and the inter-center-distance between home area and each target area was 14 cm, whereas the two target areas were separated by an inter-center distance of 15.2 cm.

The target stimulus appeared in the upper center of the screen after the cursor had spent a dwell time of 500 ms in the home area. In response to this target stimulus, participants had to move toward one of the target areas as quickly as possible (according to the target stimulus and their preceding compliance response). From this point onward, we sampled the x- and y-coordinates of the mouse cursor at 100 Hz. Initiation time (IT) was defined as the time from onset of the target stimulus until the cursor had left the home area. Movement time (MT) was recorded once the cursor hit one of the target areas; this ended the trial and the cursor shrank and disappeared from the screen. The screen was cleared 500 ms later and the next trial began after an additional 1000 ms. We did not display any error feedback when actual errors were committed but participants were encouraged to respond more quickly when they did not start their movement within 500 ms after target onset.

2.1.4. Hypotheses
For the violation group, we predicted a stronger trajectory deflection toward the opposite target (i.e., higher values for MAD and AUC) for rule violations as compared to rule-based responses. Moreover, we predicted the impact of rule compliance to be stronger in the violation group than in the reversed rule group, as should be indicated by an interaction of rule compliance and instruction group.

2.1.5. Data treatment
Trajectory data was preprocessed using custom MATLAB scripts (The MathWorks, Inc.) to determine MAD and AUC for each trial. Movements to the left were mirrored at the vertical midline, and we determined the straight line from the movement’s start point to its final point as reference. We then stripped off all dwell time data that was recorded until the cursor had left the target area (i.e., until IT), and time-normalized the remaining data to 100 points by linear interpolation. MAD was then computed as the (signed) maximum Euclidean distance from these points to the reference line (in px), with positive values indicating deviation in direction of the opposite target. Similarly, AUC was computed as the signed area between the interpolated points and reference line (in px²).

To complement the two spatial variables, we further analyzed temporal characteristics of each response in terms of the time from stimulus onset to movement initiation (initiation time; IT), and the duration of the movement (movement time; MT). For the present experimental design, effects observed on these measures should be interpreted with caution, however, because they might also indicate additional processes such as differences in response caution or distraction of attention between conditions. We therefore focused the following analyses on the more informative spatial measures of MAD and AUC.

2.2. Results
2.2.1. Trajectory analyses
For the following analyses, we omitted trials in which participants failed to act according to their compliance response (4.8%), and the immediately following trials. Furthermore, trials were
discarded as outliers if any measure (IT, MT, MAD, or AUC) deviated more than 2.5 standard deviations from the respective cell mean (6.2%). We then conducted separate split-plot analyses of variance (ANOVA) with rule compliance (correct vs. violation/reversed) as within-subjects factor and instruction group (violation vs. reversed) as between-subjects factor for each measure (see Fig. 2 and Table 1 for descriptive statistics).

The critical measures for the present question, MAD and AUC, yielded converging results. For MADS, a significant main effect of rule compliance indicated stronger deviations to the alternative target for both, violations and responses applying the reversed mapping rule, F(1,38) = 40.37, p < .001, ηp^2 = .52. Crucially, this effect was more pronounced for the violation group than for the reversed rule group, as indicated by a significant interaction of rule compliance and instruction group, F(1,38) = 11.41, p = .002, ηp^2 = .23. Overall, MADS did not differ between the two groups as the main effect of instruction group was not significant, F(1,38) = 2.67, p = .110, ηp^2 = .07. For AUCs, a similar main effect of rule compliance, F(1,38) = 31.09, p < .001, ηp^2 = .45, was moderated by a significant interaction, F(1,38) = 11.67, p < .002, ηp^2 = .23, again driven by a stronger effect for the violation group than for the reversed rule group. Overall AUCs did not differ between groups, F(1,38) = 1.79, p = .190, ηp^2 = .04.

Moreover, normal, rule-based responses were initiated more quickly than rule violations and responses that used the reversed mapping rule, as indicated by a significant main effect of rule compliance in the IT analysis, F(1,38) = 27.18, p < .001, ηp^2 = .42. 2 These effects did not differ between the groups and the overall IT level was also comparable across groups (ps > .236, ηp^2 < .04). A similar main effect of rule compliance emerged for MTS, F(1,38) = 27.68, p < .001, ηp^2 = .42. This analysis, however, also yielded an interaction of rule compliance and instruction group, F(1,38) = 4.10, p = .050, ηp^2 = .10, with a stronger impact of rule compliance for the violation group than for the reversed rule group. The violation group further showed generally longer MTS than the reversed rule group, F(1,38) = 12.17, p = .001, ηp^2 = .24.

### 2.2.2. Choice frequencies

The different labeling of the responses also affected choice frequencies: Participants in the violation group chose to violate in 39.5% of the trials whereas participants in the reversed rule group chose to use the reversed mapping rule in 47.6% of the trials. The mean proportion of violations differed from chance for the violation group, t(19) = 3.52, p = .002, Cohen’s d = 0.79, but not for the reversed rule group, t(19) = 1.38, p = .183, d = 0.31. A direct comparison of both groups also indicated that the effect on choice frequencies was larger for the violation group than for the reversed rule group, t(38) = 2.35, p = .026, d = 0.23 (corrected for unequal variances).

Furthermore, for the violation group, the frequency of violation choices was correlated with the effect of rule violations on the trajectory data. More precisely, participants with stronger effects for MAD and AUC were less likely to choose a rule violation (MAD: r = −0.41, p = .070; AUC: r = −0.45, p = .047). The same was true for MTS, r = −0.47, p = .039, but not for ITs, r = −0.09, p = .708. For the reversed rule group, none of the correlations approached significance, |r| < 0.26, ps > .285. Because the different frequencies of rule violations and reversed responses might also partly account for the reported findings, we conducted a second experiment with fixed frequencies of violations and reversed responses. In a nutshell, Experiment 2 fully replicated the trajectory data, thus ruling out this possible confound.

### 3. Experiment 2: Instructed violations

Not all violations are driven by endogenous decisions. Rather, taxonomies of rule violation behavior include “necessary violations” (Reason, 1990, 1995), i.e., rule violations that an agent has to commit for external reasons. Experiment 2 addressed such necessary violations by exchanging the freely chosen compliance response with a cue instructing the participants of the violation group whether to conform to the S–R rule (in 75% of the trials) or whether to violate it (in 25% of the trials). This setup also allows constraining the cause of the effects observed in Experiment 1 by deciding whether these effects were mainly driven by the free decision to commit a rule violation or rather driven by the labeling of the response as rule violation.

We again contrasted performance in the violation group to performance in a reversed rule group. Participants in the reversed rule group were presented with a cue that specified which mapping rule to use (Task 1 vs. Task 2, with Task 2 featuring the reversed mapping of Task 1). This procedure allowed us to assess whether the impact of rule violations would replicate even in the absence of free choices while at the same time controlling for differences in choice frequencies between both groups.

#### 3.1. Method

Experiment 2 was similar to Experiment 1 with the only exception that the compliance response was substituted for a screen that instructed the participants whether to violate a rule or not (violation group: n = 20, mean age = 26.1 years, 16 females, 1 left-handed) or whether to perform Task 1 or Task 2 (reversed rule group: n = 20, mean age = 27.4 years, 14 females, 3 left-handed, 1 ambidextrous). Participants pressed the spacebar to terminate the compliance instruction and continue with the trial proper. They completed 9 blocks, each consisting of 36 trials with correct responses and 12 trials with violations for the violation group and 36 trials with Task 1 and 12 trials with Task 2 for the reversed rule group.

#### 3.2. Results

Prior to analysis, we omitted all trials in which participants failed to act according to the compliance instructions (3.7%), and trials following such errors; 6.3% of the remaining trials were discarded as outliers due to the same criterion as for Experiment 1. The trajectory analysis did indeed replicate the main findings of Experiment 1 (Fig. 2). For MADS, a significant main effect of rule compliance again indicated stronger deviations to the alternative target for both, violations and responses applying the reversed mapping rule, F(1,38) = 24.39, p < .001, ηp^2 = .39, and this effect was again more pronounced for the violation group than for the reversed rule group, F(1,38) = 16.14, p < .001, ηp^2 = .30. Overall MADS differed between the two groups, F(1,38) = 7.44, p = .010, ηp^2 = .16. For AUCs, a similar main effect of rule compliance, F(1,38) = 21.52, p < .001, ηp^2 = .36, was moderated by a significant interaction 3. As in Experiment 1, the interaction of rule compliance and instruction group was also significant when restricting the analysis to the second half of the experiment; MADS: F(1,38) = 24.35, p < .001, ηp^2 = .39; AUCs: F(1,38) = 22.83, p < .001, ηp^2 = .35. The results again replicated when not correcting for outliers at all.
interaction, $F(1,38) = 13.54$, $p < .001$, $\eta_p^2 = .26$, again driven by a stronger effect for the violation group than for the reversed rule group. Overall AUCs differed significantly between groups, $F(1,38) = 4.87$, $p = .034$, $\eta_p^2 = .11$.

Normal, rule-based responses were also initiated more quickly than violations and responses that used the reversed mapping rule, as indicated by a significant main effect of rule compliance in the IT analysis, $F(1,38) = 57.84$, $p < .001$, $\eta_p^2 = .60$. In contrast to Experiment 1, the IT effects were larger in the violation group compared to the reversed rule group as qualified by a significant interaction, $F(1,38) = 21.12$, $p < .001$, $\eta_p^2 = .36$, whereas the overall IT level was comparable across groups, $F(1,38) = 1.90$, $p = .176$, $\eta_p^2 = .05$. Finally, a main effect of rule compliance emerged for MTs, $F(1,38) = 15.27$, $p < .001$, $\eta_p^2 = .29$, as did a significant interaction, $F(1,38) = 5.22$, $p = .028$, $\eta_p^2 = .12$, again indicating a stronger effect in the violation group than in the reversed rule group. The main effect of instruction group did not approach significance for the MT analysis, $F(1,38) = 1.01$, $p = .321$, $\eta_p^2 = .03$.

4. Discussion
The reported findings show that rules are not easily broken; and even if they are broken eventually, behavior is still attracted toward compliance. Furthermore, it is not the free decision to violate that drives this effect but rather the mere fact of labeling a behavior as rule violation. Merely defining a rule, however arbitrary and irrelevant it may be, thus seems to prompt a tendency toward following it. Accordingly, rule compliance in humans might be a behavioral default that arises at least partly as a consequence of the cognitive burdens of rule violations, even without any ethical or moral implications.
The continued impact of the rule representation on violation behavior therefore reminds of the behavioral signature of the rule representation on violation behavior itself. Finally, open questions relate to the impact of rules and norms in the context of social interactions. In these terms, previous research has shown a remarkable sensitivity to implicit norms such as obeying to social requests (Sartori, Becchio, Bulgheroni, & Castelli, 2009), suggesting that rule violations might leave an even more pronounced fingerprint on the agent’s behavior in such settings. In any case, the reported findings call for a new perspective on rule violation behavior – a perspective that focuses on the agent violating a rule and the processes involved in actually performing the behavior.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.cognition.2015.11.009.

References


This analysis and interpretation was suggested by an anonymous reviewer.