Psychological Research (2016) 80:838–852
DOI 10.1007/s00426-015-0690-9

ORIGINAL ARTICLE

Pushing the rules: effects and aftereffects of deliberate rule violations

Robert Wirth[1] · **Roland Pfister**[1] · **Anna Foerster**[1] · **Lynn Huestegge**[1] · **Wilfried Kunde**[1]

Received: 9 January 2015 / Accepted: 15 July 2015 / Published online: 6 August 2015

Abstract Most of our daily life is organized around rules and social norms. But what makes rules so special? And what if one were to break a rule intentionally? Can we simply free us from the present set of rules or do we automatically adhere to them? How do rule violations influence subsequent behavior? To investigate the effects and aftereffects of violating simple S-R rule, we conducted three experiments that investigated continuous finger-tracking responses on an iPad. Our experiments show that rule violations are distinct from rule-based actions in both response times and movement trajectories, they take longer to initiate and execute, and their movement trajectory is heavily contorted. Data not only show differences between the two types of response (rule-based vs. violation), but also yielded a characteristic pattern of aftereffects in case of rule violations: rule violations do not trigger adaptation effects that render further rule violations less difficult, but every rule violation poses repeated effort on the agent. The study represents a first step towards understanding the signature and underlying mechanisms of deliberate rule violations, they cannot be acted out by themselves, but require the activation of the original rule first. Consequently, they are best understood as reformulations of existing rules that are not accessible on their own, but need to be constantly derived from the original rule, with an add-on that might entail an active tendency to steer away from mental representations that reflect (socially) unwanted behavior.

Introduction

What if we lived in a world without rules, or a world without punishment for violating moral codes? In 1974, performance artist Marina Abramović challenged this question in her famous work “Rhythm 0”. She placed a range of objects on a table, e.g., a feather, a rose, a gun, and encouraged the audience to do whatever they wanted to do to her—freed from all responsibility. The audience played along, but shy at first. Eventually they kissed her, cut her clothes, carried her around, and even aimed the gun at her. During the entire performance, she did not respond and stood still as if she was a mere object. But at the end of the performance, when she started moving again, the audience was overwhelmed by what they had done to the artist, they could not face her any longer and took flight from the performance.

It seems as if it is challenging to manage situations in which familiar rules do not apply. The case of “Rhythm 0” certainly documents a rather special and extreme situation that challenged normative moral standards. Rules, however, can be defined on a wide spectrum (Reason, 1990, 1995), ranging from such moral norms on the one end to simple instructions about what to do on the other end (Pfister, 2013). Whereas it is clear that the former type of rules has a profound impact on human behavior (Asch, 1956; Milgram, 1963), it is not clear whether and how merely instructed rules would also affect behavior of agents who violate them. With the present experiments, we therefore addressed whether there is something special about violating simple S-R rules that were set up by instructions. Can we free ourselves from the present set of rules in this case or do we still have a tendency to adhere to them? Are violations of simple S-R rules accompanied by a distinct behavioral signature?

✉ Robert Wirth
robert.wirth@uni-wuerzburg.de

[1] Department of Psychology, Julius-Maximilians-University of Würzburg, Röntgenring 11, 97070 Würzburg, Germany

Initial evidence suggests that it is indeed hard to overcome even simple, arbitrary rules (Pfister, Wirth, Schwarz, Steinhauser, & Kunde, submitted): when violating a rule, the original rule remains activated and therefore shapes our behavior. In these experiments, an impact of the rule representation was visible in terms of movement trajectories that were attracted toward the rule-conform option during rule violation. It is almost ironic that when we try hard not to follow a rule, this is exactly when we cannot suppress its influence (Wegner, 2009). Even though these findings are suggestive of a continued rule representation during violation behavior, they can merely represent a first step towards understanding the cognitive architecture of deliberate rule violations.

One critical limitation of the described work is its focus on rule violations as single, isolated instances. With the present experiments, we aimed at setting rule violations in context by investigating the impact of previous instances of rule-based or violation behavior and the impact of different instructional framings on an agent's performance. Based on the assumption that rule violations entail a conflict between the rule-based and the violation response, we assume to find conflict adaptation processes (Botvinick, Barch, Carter, & Cohen, 2001; Gratton, Coles, & Donchin, 1992). We therefore adapted previous methods (Pfister et al., submitted) to investigate which cognitive processes go along with rule violations, and, importantly, how these violations influence subsequent behavior. These experiments were designed to capture and compare parameters of not only the decision process between rule-based and violation behavior, but also of the execution of the response. Namely, we analyzed the movement trajectories of participants' sweeping responses on the touchscreen of an iPad while they followed or violated an instructed rule, which required the movement of the finger to a certain location, according to a certain stimulus. Based on these trajectories, we can compute specific parameters that mirror specific cognitive processes, i.e., the speed of response planning, or spatial and temporal aspects of the response execution (Pfister, Janczyk, Wirth, Dignath, & Kunde, 2014; Wirth, Pfister, & Kunde, 2015; Wirth et al., 2015). Experiment 1 employed one simple rule that had to be violated at times, whereas Experiment 2 added a second rule that specifically called for the response that was previously labeled "violation", to control for effects of responding in a reversed rule mapping (Schroder, Moran, Moser, & Altmann, 2012). Finally, Experiment 3 provided an additional control group by addressing inversions of an instructed rule as compared to two reversed rules in Experiment 2. A direct comparison of the experiments therefore allowed us to pinpoint specific effects and aftereffects of violating a simple S-R rule.

Experiment 1

Introduction

Experiment 1 was designed (a) to quantify the difficulty that violations pose on the acting agent and (b) to investigate the impact of such violations on subsequent behavior. We used a simple S-R rule that mapped two target stimuli to a left and a right sweeping response on an iPad. This rule had to be followed most of the time, but had to be violated in a fraction of trials (i.e., akin to the definition of "necessary violations"; Reason, 1990, 1995). We applied a two-dimensional finger-tracking design to not only depict the impact of violations in terms of an extra amount of processing time, but also in terms of distinct spatial signatures. Participants had to sweep their finger from a starting area in the bottom center of the display to an upper-left or an upper-right target area on the iPad's touchscreen. The critical question was whether the movement trajectories would vary as a function of current response type (rule-based vs. violation behavior) and, crucially, also as a function of preceding response type.

Methods

Participants

Twenty participants were recruited (mean age = 21.0 years, SD = 2.3, 5 male, 3 left-handed) and received either course credit or €5 monetary compensation. All participants gave informed consent, were naïve to the purpose of the experiment and were debriefed after the session.

Apparatus and stimuli

The experiment was run on an iPad in portrait mode, which sampled the participants' finger movements at 100 Hz. Viewing distance was about 50 cm. We used two chess symbols (king, ♔, and pawn, ♙) as target stimuli to prompt movements to the left or to the right target area (two circles of 2 cm in diameter in the upper left and right corners of the display). The target areas were separated by 11 cm. In between trials, the two chess symbols in the center of the screen reminded participants which symbol called for a movement to the left (the one displayed on the left side) and which symbol called for a movement to the right (the one displayed on the right side). A written instruction between the two chess figures instructed the rule compliance for the following trial. The starting position for the movement (1 cm in diameter) was located at the bottom center of the screen, 17 cm from the middle of the two target positions at an angle of 31° to each side. Stimuli were presented against a light gray background (see Fig. 1).

Procedure

Participants started each trial by touching the starting area with the index finger of the dominant hand. Immediately, a target symbol appeared in the center of the screen to indicate whether a movement to the left or a movement to the right had to be executed. Simultaneously, the reminder of the S-R mapping and the written instruction disappeared. Half of the participants were instructed to make a smooth finger movement to the left target area if the center showed a pawn symbol and to the right target area if it showed a king symbol. The other half of the participants was instructed with the opposite S-R mapping for counterbalancing. The target symbol disappeared as soon as the finger left the starting area. One out of four trials included a written instruction to rule violation instead of rule-based behavior before trial start (for example "♚ break the rule ♙", displayed in between trials). In these trials, the displayed mapping rule had to be violated; the response that a target required originally was now contraindicated. A trial ended when the finger was lifted from the touchscreen. Error feedback was displayed only if participants failed to hit one of the designated target areas. Participants were instructed to respond quickly and accurately; still the experiment was self-paced, so participants chose on their own when to start a trial and how long they took breaks in between blocks. Participants completed 12 blocks of 48 trials, with each of the target symbols presented equally often.

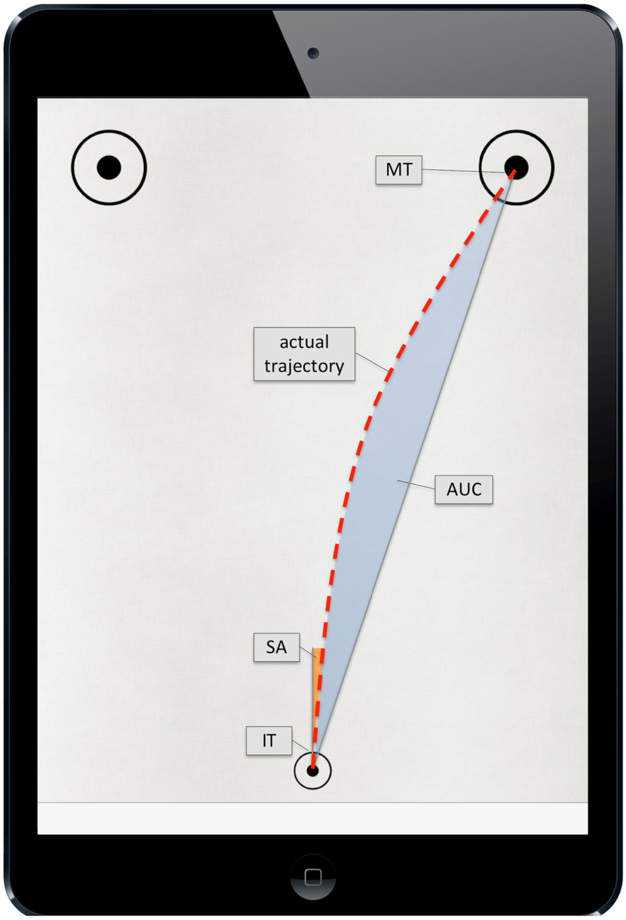

Fig. 1 Setting of the experiments and relevant measures. Participants dragged their finger in a continuous movement from the starting area on the *bottom* of the screen to one of the two areas in the *upper corners* of the screen. In Experiment 1, they followed an instructed mapping rule in 75 % of the trials whereas they violated the rule in 25 % of the trials. In Experiment 2, participants performed an instructed primary task in 75 % of the trials and performed a task with reversed mapping in 25 % of the trials. In Experiment 3, participants followed an instructed mapping rule in 75 % of the trials and had to invert the rule in 25 % of the trials. Initiation time (*IT*) was defined as the time from target onset to movement initiation, movement time (*MT*) as the time of movement execution. Starting angle (*SA*) mirrors the angle of the movement trajectory against the *vertical midline* (*orange*) upon leaving the home area; area under the curve (*AUC*) measures the area between the actual trajectory and a *straight line* from start- to endpoint (*blue*) (color figure online)

Results

Preprocessing

We analyzed four variables of each movement: the time from stimulus onset to movement initiation (initiation time; IT), the duration of the movement (movement time; MT), the angle between the trajectory and the vertical midline at response initiation (starting angle; SA) and the area between the actual trajectory and a straight line from start- to endpoint (area under the curve; AUC). IT therefore mirrors the speed of response selection and motor planning; MT, SA and AUC depict specific temporal and spatial parameters of the executed response. Positive values for AUC and smaller (or negative) values of SA indicate that a movement is attracted to the competing response alternative, indicating a persisting influence of the original mapping rule.

IT was defined as the time that it takes for the finger to leave the starting area. From this point, x- and y-coordinates were recorded; MT was determined when the finger left the touchscreen. AUC and SA were computed from the time-normalized coordinate data of each trial using custom MATLAB scripts (The Mathworks, Inc.). Movements to the left were mirrored at the vertical midline for all analyses. AUC was computed as the signed area relative to a straight line from start- to endpoint of the movement (positive values indicating attraction toward the opposite side, negative values indicating attraction toward the nearest edge of the display). SA was defined as the angle between the actual trajectory and the vertical midline (see Fig. 1, negative values indicating attraction toward the opposite side, positive values indicating attraction toward the rule-based target area).

Data selection and analyses

For the following analyses, we omitted trials in which participants failed to act according to the instruction (3.5 %) and the immediately following trials (3.0 %). We also excluded trials in which participants failed to hit any of the two target areas at all (2.5 %). Trials were discarded as outliers if any of the measures (IT, MT, SA, AUC) deviated more than 2.5 standard deviations from the respective cell mean (6.3 %). Each measure was then analyzed in a separate 2 × 2 ANOVA with current response type (rule-based vs. violation) and preceding response type as within-subject factors (see Fig. 2). Additionally, repetition benefits for each measure and each response type were computed as the difference between switch and repetition trials. That is, repetition benefit for rule-based responses in IT was computed as (IT of rule-based responses after violation trial) minus (IT of rule-based responses after rule-based trial); all other repetition benefits were computed accordingly. Repetition benefits are only mentioned if significant.

Initiation times

A significant effect of current response type, $F(1,19) = 26.14$, $p < .001$, $\eta_p^2 = .58$, was driven by slower response initiation for violations (450 ms) than for rule-based behavior (392 ms). A similar effect emerged for preceding response type, $F(1,19) = 10.43$, $p = .004$, $\eta_p^2 = .35$, with slower responses following violations (428 ms) compared to rule-based behavior (399 ms). The interaction between preceding and current response type was also significant, $F(1,19) = 32.34$, $p < .001$, $\eta_p^2 = .63$, with a profound effect of current response type only after rule-based responses ($\Delta = 73$ ms), $t(19) = 6.04$, $p < .001$, $d = 1.35$, but not after violation responses ($\Delta = 10$ ms), $t(19) = 1.47$, $p = .157$, $d = 0.33$. Repetition benefits were smaller for violation responses ($\Delta = 20$ ms), $t(19) = 3.54$, $p = .002$, $d = 0.79$, compared to rule-based responses ($\Delta = 43$ ms), $t(19) = 5.82$, $p = .002$, $d = 1.30$ (Fig. 2a).

Starting angles

A significant effect of current response type, $F(1,19) = 27.18$, $p < .001$, $\eta_p^2 = .59$, indicated shallower initial trajectories for rule-based behavior (1.6°) compared to violations, which were steeper and initially directed to the opposite side (−2.2°). Neither preceding response type nor the interaction approached significance, $F\text{s} < 1$ (Fig. 2b).

Movement times

Response execution was slower for violations (628 ms) than for rule-based behavior (581 ms), $F(1,19) = 29.52$, $p < .001$, $\eta_p^2 = .61$. A significant effect of preceding response type, $F(1,19) = 10.84$, $p = .004$, $\eta_p^2 = .36$, further indicated slower movements following violations (608 ms) compared to rule-based behavior (588 ms). The interaction between preceding and current response type was not significant, $F(1,19) = 1.48$, $p = .239$, $\eta_p^2 = .07$. Rule-based responses produced repetition benefits ($\Delta = 12$ ms), $t(19) = 2.19$, $p = .041$, $d = 0.49$, while violation responses produced a negative repetition benefit (repetition costs), with repeated violations leading to slower movements compared to single instances ($\Delta = -34$ ms), $t(19) = 2.28$, $p = .034$, $d = 0.51$ (Fig. 2c).

Areas under the curve

A significant effect for current response type, $F(1,19) = 46.56$, $p < .001$, $\eta_p^2 = .71$, again indicated more curved trajectories for violations (45073px^2) than for rule-based behavior (28464px^2). The effect of preceding response type was marginally significant, $F(1,19) = 4.24$, $p = .053$, $\eta_p^2 = .18$, with descriptively more curved trajectories following violations (36359px^2) compared to rule-based behavior (31183px^2). The interaction between the two factors did not approach significance, $F < 1$. Repetition benefits were significant only for rule-based responses ($\Delta = 4891\text{px}^2$), $t(19) = 3.86$, $p = .001$, $d = 0.86$ (Fig. 2d).

Discussion

In Experiment 1, we investigated the difficulties that rule violations pose on the acting agent. Replicating previous findings (Pfister et al., submitted), we found violation responses to be more effortful than rule-based responses. They took longer to initiate and execute, and their movement trajectory is heavily deflected towards the alternative target, suggestive of a continued influence of the original mapping rule.

The resulting pattern of ITs further suggests that repeatedly violating a rule facilitates the initiation of rule violations. This finding reminds of sequential patterns that are typically reported by studies on cognitive conflict and conflict adaptation (Botvinick et al., 2001; Gratton et al., 1992). This could ultimately suggest that the planning of a violation response is associated with cognitive conflict between the automatic rule-based and the violation response, and that this conflict lessens with previous

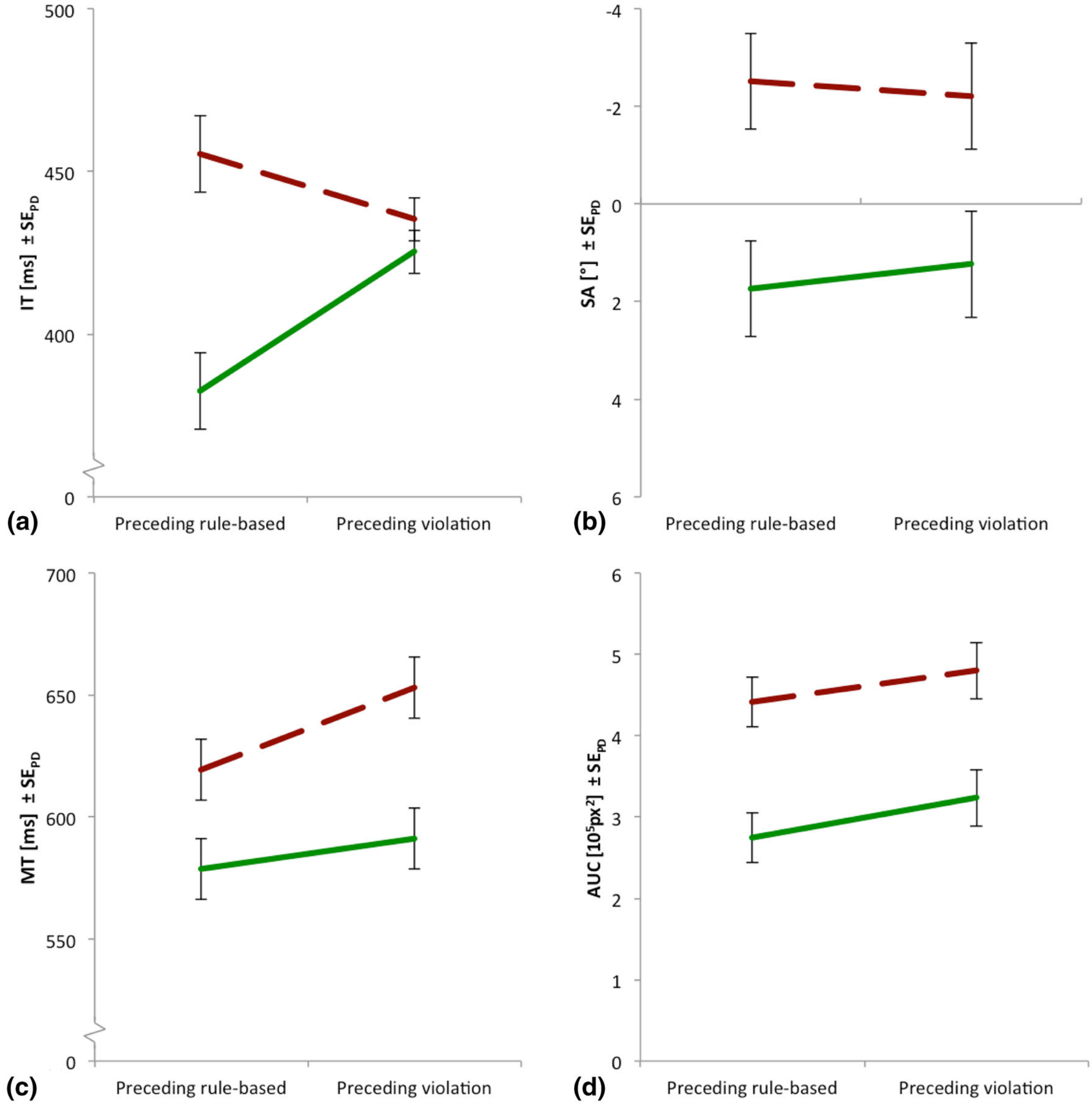

Fig. 2 Results for Experiment 1 (violation instructions). Initiation times (*ITs*; **a**), starting angles (*SA*; **b**), movement times (*MT*; **c**) and areas under the curve (*AUC*; **d**) are plotted as a function of preceding response type (abscissa) and current response type (*continuous line* for rule-based responses; *dashed line* for violation responses). *Error bars* represent standard errors of paired differences, calculated separately for each instance of preceding response type (Pfister & Janczyk, 2013)

violation responses (for a complementary explanation in terms of task-switching, Monsell, 2003, see the “General discussion”).

Surprisingly, however, there were no sequential effects of rule violations on the actual movement trajectories. That is, the signature of rule violations on SA, MT, and AUC remained visible even after having committed a rule violation only a few seconds before. Participants thus appear not to adjust their response execution according to recent events after a rule violation. Before drawing further conclusions from these findings, two experiments provide important control conditions to clarify the interpretation of these data.

Experiment 2

Introduction

Experiment 2 investigated whether the pattern of results observed in the preceding experiment is specific to rule violations or whether they represent just an instance of task switching (Monsell, 2003). We did this by employing the same task as in Experiment 1, but slightly varied the instructions. Instead of prompting participants to follow or break a given rule, we introduced two response mappings that were labeled "Task 1" and "Task 2", with Task 2 being the reversed mapping of Task 1 (Schroder et al., 2012). As Task 2 was presented equally often as the violation prompt in Experiment 1, participants virtually had to employ the exact same responses in both experiments. In Experiment 2, however, participants were presented with two equally neutral and separate task sets, whereas the corresponding actions were labeled as deviant behavior in Experiment 1.

Methods

Participants

A new set of twenty participants was recruited (mean age = 21.8 years, SD = 4.2, 4 male, 2 left-handed) and received either course credit or €5 monetary compensation. All participants gave informed consent, were naïve to the purpose of the experiment and were debriefed after the session.

Apparatus, stimuli and procedure

The experiment was mostly identical to the first experiment. But instead of instructing participants to break a given rule in one out of four trials, participants were asked to complete either "Task 1" (frequent task) or "Task 2" (infrequent task), with Task 2 consisting of the inverted S-R mapping of Task 1 and occurring in one out of four trials. This way, Experiment 2 required the exact same movements as Experiment 1, but instead of introducing rule-based and violation behavior, participants were presented with two separate and equally neutral response mappings.

Results

Data treatment and analyses

The data were treated exactly as in Experiment 1. We omitted trials in which participants failed to act according to the instructions (3.5 %), the immediately following trials (3.0 %) and trials in which participants failed to hit any of the two target areas (2.9 %). Trials were discarded as outliers if any of the measures (IT, MT, SA, AUC) deviated more than 2.5 standard deviations from the respective cell mean (6.8 %). The four measures were then analyzed in separate 2 × 2 ANOVAs with current response type (frequent task vs. infrequent task) and preceding response type as within-subject factors (Fig. 3).

Initiation times

A significant effect of current response type, $F(1,19) = 6.82$, $p = .017$, $\eta_p^2 = .26$, indicated slower response initiation for the infrequent task (472 ms) than for the frequent task (442 ms). The interaction between preceding response type and current response type was also significant, $F(1,19) = 11.84$, $p = .003$, $\eta_p^2 = .38$, with a pronounced effect of response type after frequent tasks ($\Delta = 40$ ms), $t(19) = 3.29$, $p = .004$, $d = 0.73$, and no response costs after infrequent tasks ($\Delta = 0$ ms), $t(19) = 0.12$, $p = .903$, $d = 0.03$. Repetition benefits were significant for the frequent task ($\Delta = 24$ ms), $t(19) = 2.74$, $p = .013$, $d = 0.61$, and marginally significant for the infrequent task ($\Delta = 16$ ms), $t(19) = 1.86$, $p = .079$, $d = 0.42$ (Fig. 3a).

Starting angles

A significant effect of current response type, $F(1,19) = 7.06$, $p = .016$, $\eta_p^2 = .27$, was driven by shallower response initiation for the frequent task (3.7°) compared to the infrequent task (0.2°). A similar effect of preceding response type emerged, $F(1,19) = 12.67$, $p = .002$, $\eta_p^2 = .40$, with shallower responses following infrequent tasks (3.4°) compared to frequent tasks (2.8°). The interaction between preceding response type and current response type was also significant, $F(1,19) = 17.58$, $p < .001$, $\eta_p^2 = .48$, with effects of response type after frequent tasks ($\Delta = -5.1°$), $t(19) = -4.07$, $p = .001$, $d = 0.91$, and no significant differences after infrequent tasks ($\Delta = 0.7°$), $t(19) = 0.90$, $p = .328$, $d = 0.18$. Repetition benefits were only significant for the infrequent task ($\Delta = 5.2°$), $t(19) = 5.01$, $p = .002$, $d = 1.12$ (Fig. 3b).

Movement times

A significant effect of current response type emerged, $F(1,19) = 9.20$, $p = .007$, $\eta_p^2 = .33$, with slower movements on infrequent tasks (576 ms) than on frequent tasks (552 ms), as well as a significant effect of preceding response type, $F(1,19) = 10.17$, $p = .005$, $\eta_p^2 = .35$, with slightly faster movements following infrequent tasks (556 ms) compared to frequent tasks (568 ms). The interaction between preceding response type and current response type was also significant, $F(1,19) = 18.85$, $p < .001$, $\eta_p^2 = .50$, indicating a pronounced effect of response type after frequent tasks ($\Delta = 34$ ms), $t(19) = 4.23$, $p < .001$, $d = 0.95$, and no response costs after

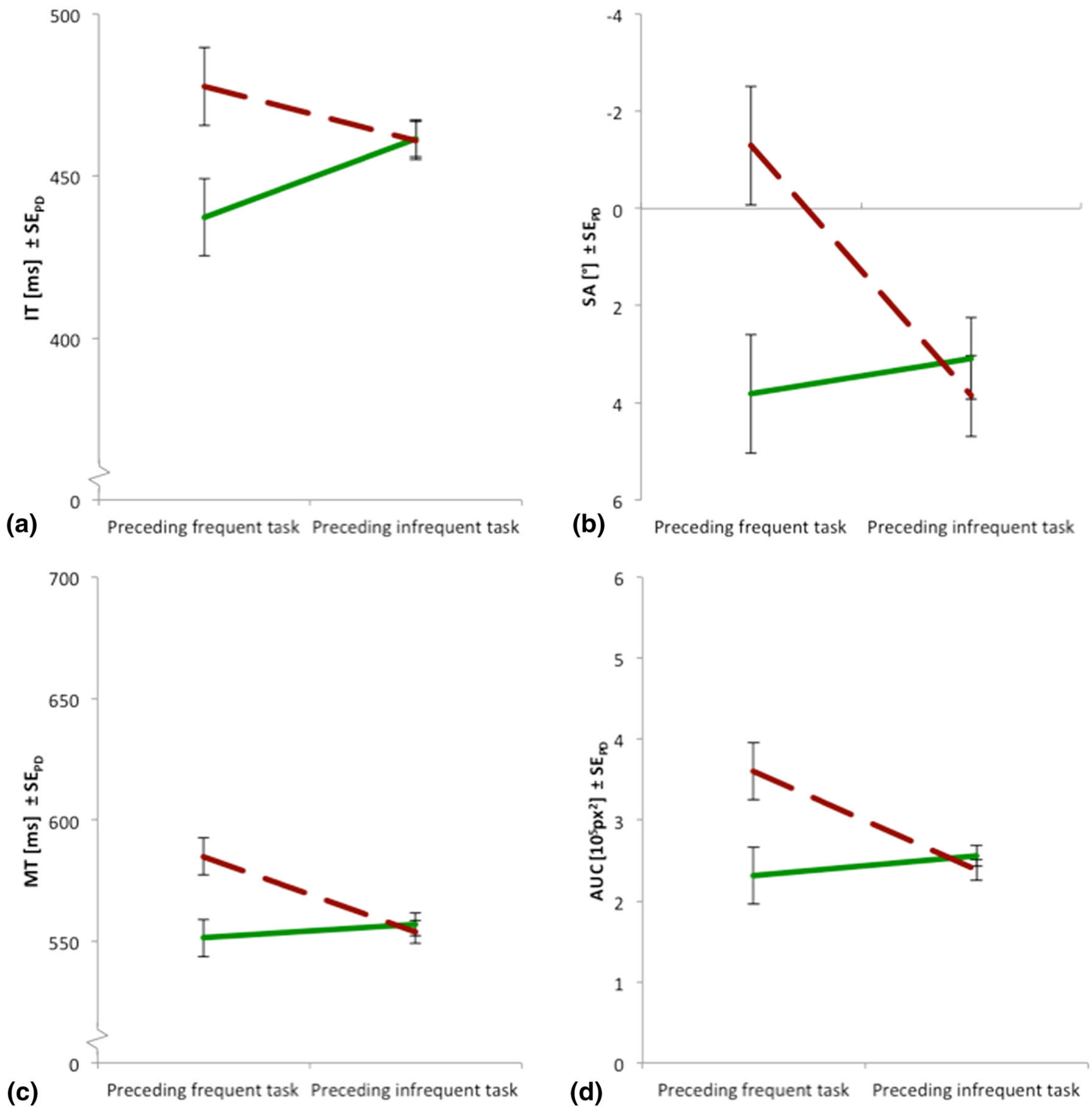

Fig. 3 Results for Experiment 2 (task-switching instructions). Initiation times (*ITs*; **a**), starting angles (*SA*; **b**), movement times (*MT*; **c**) and areas under the curve (*AUC*; **d**) are plotted as a function of preceding response type (abscissa) and current response type (*continuous line* for responses to the frequent task; *dashed line* for responses to the infrequent task). *Error bars* represent standard errors of paired differences, calculated separately for each instance of preceding response type (Pfister & Janczyk, 2013)

infrequent tasks ($\Delta = -3$ ms), $t(19) = -0.65$, $p = .524$, $d = 0.14$. Repetition benefits were significant for infrequent tasks ($\Delta = 31$ ms), $t(19) = 4.08$, $p = .001$, $d = 0.91$, and marginally significant for frequent tasks ($\Delta = 5$ ms), $t(19) = 1.79$, $p = .090$, $d = 0.40$ (Fig. 3c).

Areas under the curve

A significant effect for current response type, $F(1,19) = 8.10$, $p = .010$, $\eta_p^2 = .30$, was driven by more curved response execution on infrequent tasks (32464px^2) than on frequent tasks (23571px^2). A similar effect of preceding response type emerged, $F(1,19) = 9.45$, $p = .006$, $\eta_p^2 = .33$, with less curved response execution following infrequent tasks (24989px^2) compared to frequent tasks (25705px^2). The interaction between preceding response type and current response type was also significant, $F(1,19) = 15.22$, $p = .001$, $\eta_p^2 = .45$, with bigger effects of response type after frequent tasks ($\Delta = 12873$px^2), $t(19) = 3.59$, $p < .001$, $d = 0.80$, and descriptively reversed response costs after infrequent tasks ($\Delta = -1757$px^2), $t(19) = -1.31$, $p = .204$, $d = 0.29$. Repetition benefits were significant for infrequent tasks ($\Delta = 12154$px^2), $t(19) = 3.78$, $p = .001$, $d = 0.85$, and marginally significant for frequent tasks ($\Delta = 2477\text{px}^2$), $t(19) = 1.93$, $p = .069$, $d = 0.43$ (Fig. 3d).

Discussion

In Experiment 2, we slightly changed the task instructions, as compared to Experiment 1: instead of instructing one task set that had to be violated, we provided participants with two separate task sets that called for the exact same behavior as in Experiment 1.

We found that infrequent, rule-based behavior still differs from frequent, rule-based behavior with infrequent behavior being more effortful in both, planning and execution. Reversed behavior is also deflected to the opposite side, which could indicate an influence of the dominant rule of Task 1, or reflect task-switching effects between the two instructed task sets (Monsell, 2003, see the "General discussion" for a more thorough discussion). As of now, we can conclude that task-switching effects and the presentation frequencies that we used here could potentially account for the signature that we found to be associated with rule violations. However, the effects observed in Experiment 2 were substantially smaller than those observed in Experiment 1 and came with a different pattern of adaptations according to recent events (for a corresponding between-experiment analysis, see "Between-experiment analyses"). These diverging results might be driven by two procedural differences: the labeling of the infrequent response as rule violation vs. an alternative but rule-conform option for one, and the instruction in terms of one vs. two task sets for another. Both differences might partly account for these diverging results and Experiment 3 therefore aimed at clarifying the role of both contributions.

Experiment 3

Introduction

In Experiment 2, participants were instructed with two separate task sets for both conditions (frequent vs. infrequent task), whereas Experiment 1 only employed one task set (rule-based), while rule violations had to be derived from the instructed one. To test whether this difference in available task sets can account for the specific behavioral signatures of rule violations compared to task switches, Experiment 3 provided an additional control condition that only provided one task set, and participants had to derive the alternative responses from this task set by inversion (Wason, 1959; Wegner, 2009). But compared to Experiment 1, we now employed an instruction that put less emphasis on the deviating nature of the infrequent task, but offered a more neutral response alternative.

Methods

Participants

A new set of twenty participants was recruited (mean age = 23.4 years, SD = 2.6, 4 male, 3 left-handed) and received either course credit or €5 monetary compensation. All participants gave informed consent, were naïve to the purpose of the experiment and were debriefed after the session.

Apparatus, stimuli and procedure

The experiment was mostly identical to the first experiment. But instead of instructing participants to break a given rule in one out of four trials, participants were asked to either "follow the standard rule" or "invert the rule". This way, Experiment 3 required the exact same movements as Experiments 1 and 2, but, as in Experiment 1, now only instructed one task set. At the same time, we took care to instruct the inversion as part of the mapping rule rather than labeling the behavior as violation as we had done in Experiment 1.

Results

Data treatment and analyses

The data were treated exactly as in Experiments 1 and 2. We omitted trials in which participants failed to act according to the instructions (3.4 %), the immediately following trials (3.3 %) and trials in which participants failed to hit any of the two target areas (2.3 %). Trials were discarded as outliers if any of the measures (IT, MT, SA, AUC) deviated more than 2.5 standard deviations from the respective cell mean (6.3 %). The four measures were then analyzed in separate 2 × 2 ANOVAs with current response type (standard vs. inverted) and preceding response type as within-subject factors (Fig. 4).

Initiation times

A significant effect of current response type, $F(1,19) = 14.68$, $p = .001$, $\eta_p^2 = .43$, was driven by slower response initiation for inversions (411 ms) than for standard responses (387 ms). A similar effect emerged for preceding response type, $F(1,19) = 5.08$, $p = .036$, $\eta_p^2 = .21$, with slower responses following inversions (404 ms) compared to standard responses (395 ms). The interaction between preceding and current response type was also significant, $F(1,19) = 17.04$, $p = .001$, $\eta_p^2 = .47$, with a profound effect of current response type only after

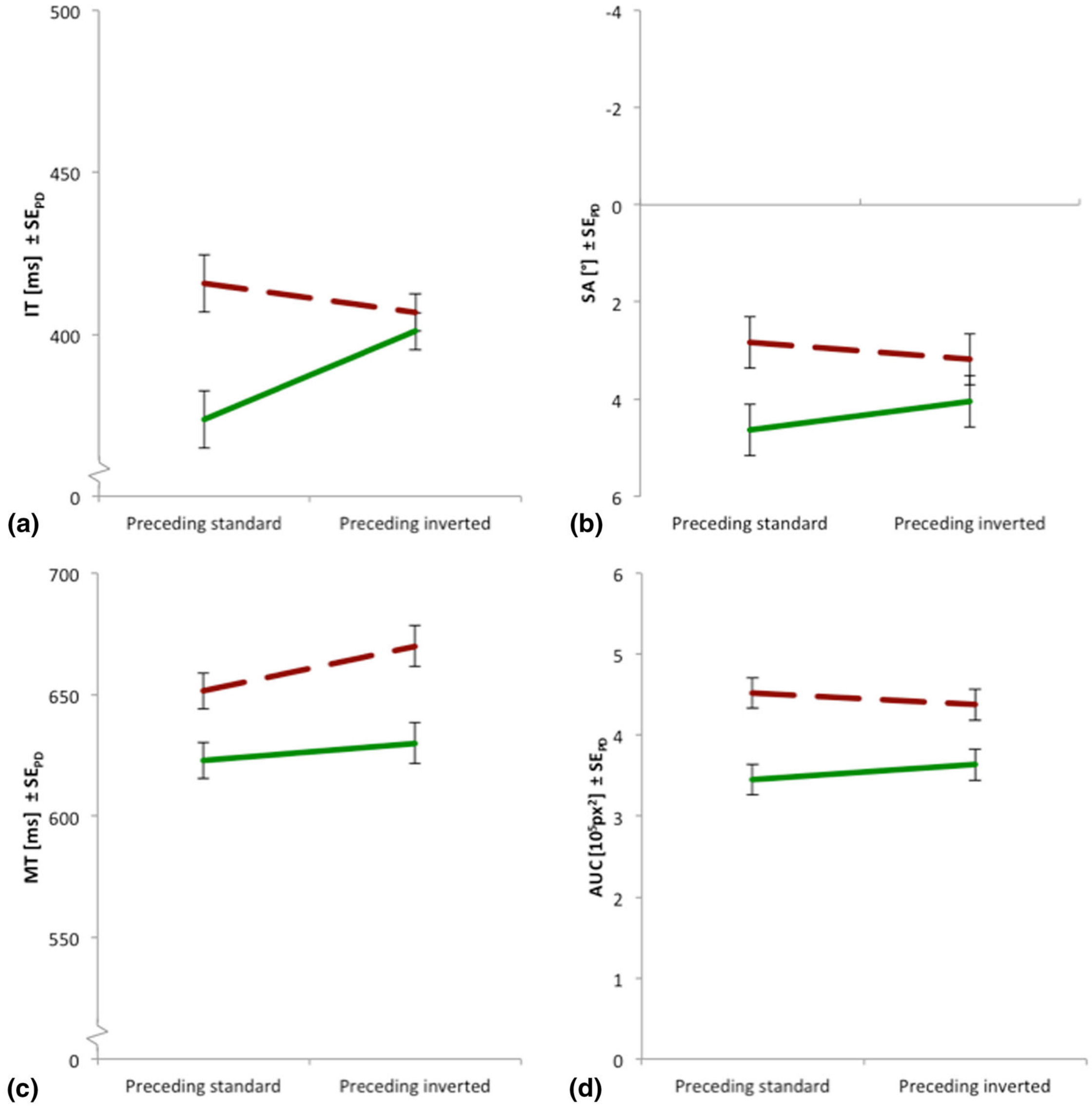

Fig. 4 Results for Experiment 3 (inversion instructions). Initiation times (*ITs*; **a**), starting angles (*SA*; **b**), movement times (*MT*; **c**) and areas under the curve (*AUC*; **d**) are plotted as a function of preceding response type (abscissa) and current response type (*continuous line* for standard responses according to the original rule; *dashed line* for inverted responses). *Error bars* represent standard errors of paired differences, calculated separately for each instance of preceding response type (Pfister & Janczyk, 2013)

standard responses ($\Delta = 42$ ms), $t(19) = 4.65$, $p < .001$, $d = 1.04$, but not after inverted responses ($\Delta = 6$ ms), $t(19) = 0.98$, $p = .339$, $d = 0.22$. Repetition benefits were significant only for standard responses ($\Delta = 27$ ms), $t(19) = 4.77$, $p < .001$, $d = 1.07$ (Fig. 4a).

Starting angles

A significant effect of current response type, $F(1,19) = 13.07$, $p = .002$, $\eta_p^2 = .41$, indicated shallower initial trajectories for standard responses (4.4°) compared to inversions (3.0°). Neither preceding response type nor the interaction approached significance, $Fs < 1.39$, $ps > .254$ (Fig. 4b).

Movement times

Response execution was slower for inversions (661 ms) than for standard responses (627 ms), $F(1,19) = 23.85$, $p < .001$, $\eta_p^2 = .56$. A significant effect of preceding response type, $F(1,19) = 7.96$, $p = .011$, $\eta_p^2 = .30$, further indicated slower movements following inversions (650 ms) compared to standard responses (637 ms). The interaction between preceding and current response type was not significant, $F(1,19) = 1.91$, $p = .183$, $\eta_p^2 = .09$. Unexpectedly, inverted responses produced repetition costs rather than benefits ($\Delta = -18$ ms), $t(19) = 2.74$, $p = .034$, $d = 0.61$ (Fig. 4c).

Areas under the curve

A significant effect for current response type, $F(1,19) = 33.69$, $p < .001$, $\eta_p^2 = .64$, again indicated more curved trajectories for inverted responses (44494px^2) compared to standard responses (35423px^2). No other effects reached significance, *F*s < 2.02, *p*s > .172. Standard responses produced marginally significant repetition benefits ($\Delta = 1850$px^2), $t(19) = 2.88$, $p = .075$, $d = 0.42$ (Fig. 4d).

Discussion

In Experiment 3, we tested whether the instruction of a single task set that had to be inverted could account for the pattern of data that we found for violation responses in Experiment 1. And indeed, we again found that responses based on the inverted task set were slower and more attracted to the competing response alternative. Moreover, we were able to replicate the sequential adaptation effect of ITs and the additive effect of SAs, MTs and AUCs that indicate that the selection and planning of an inverted response becomes more efficient with previous experience, while the execution of these responses does not.

To compare the size of the response costs and adaptation effects that come with rule violations compared to task switches and inversions, we conducted between-experiment analyses.

Between-experiment analyses

Results

An overview of the differences and similarities between the three experiments is listed in Table 1. This overview includes both differences regarding the experimental design and regarding the main results.

For all between-experiment analyses, we conducted ANOVAs on the immediate effects of the experimental manipulation and on the corresponding sequential effects. Immediate effects were computed as the mean differences between the two current response types (violation/infrequent/inverted minus rule-based/frequent/standard) and they were analyzed by a one-way ANOVA with experiment (Exp. 1: violation instructions vs. Exp. 2: task-switching instructions vs. Exp.3: inversion instructions) as between-subjects factor. To reduce redundancy, we focus on planned contrasts that pitted Experiment 1 against Experiment 2, and Experiment 1 against Experiment 3, and we finally tested the contrast between Experiments 2 and 3.

Sequential effects were computed as the differences between the effects after the two response types (violation-/infrequency-/inversion-effect after rule-based/frequent/standard responses minus effect after deviant responses) and they were analyzed by a one-way ANOVA with experiment (Exp. 1: violation instructions vs. Exp. 2: task-switching instructions vs. Exp.3: inversion instructions) as between-subjects factor. Then, we focused on planned contrasts that compared the adaptation effect between Experiment 1 against Experiment 2, and those that compared the adaptation effect between Experiment 1 and Experiment 3. We finally compared the adaptation between Experiments 2 and 3. Because we expected the effects of rule violations (Exp. 1) to exceed the effects of inversion (Exp. 3) and, likewise the effects of inversion to exceed the effects of task frequency (Exp. 2), we report the following contrast estimates as one-tailed.

Initiation times

Regarding the immediate effects, the comparison of the effect size of Experiment 1 ($\Delta = 40$ ms) against Experiment 2 ($\Delta = 19$ ms) was significant, $t_{1/2}(57) = 2.10$, $p = .020$, $d = 0.63$, while the contrast between Experiment 1 and Experiment 3 ($\Delta = 24$ ms) only produced a marginally significant effect, $t_{1/3}(57) = 1.58$, $p = .060$, $d = 0.51$. The comparison between Experiments 2 and 3 was not significant, $t_{2/3} < 1$.

For sequential effects, only the comparison of Experiment 1 ($\Delta = 60$ ms) against Experiment 3 ($\Delta = 36$ ms) was marginally significant, $t_{1/3}(57) = 1.42$, $p = .085$, $d = 0.52$, the adaptation effect in Experiment 2 ($\Delta = 40$ ms) differed from neither experiment, *t*s < 1.17, *p*s > .123.

Starting angles

Regarding the immediate effects, the comparison of the effect size of Experiment 1 ($\Delta = -4.27°$) against Experiment 2 ($\Delta = -1.34°$) was significant, $t_{1/2}(57) = 1.78$, $p = .041$, $d = 0.86$, as was the contrast between Experiment 1 and Experiment 3 ($\Delta = -2.43°$), $t_{1/3}(57) = 2.83$, $p = .003$, $d = 0.71$. The comparison between Experiments 2 and 3 was not significant, $t_{2/3}(57) = 1.05$, $p = .144$, $d = 0.38$.

For sequential effects, the comparison of Experiment 1 ($\Delta = -0.88°$) against Experiment 2 ($\Delta = -5.09°$) was significant, $t_{1/2}(57) = 2.57$, $p = .007$, $d = 0.74$, as was the contrast between Experiment 2 and Experiment 3 ($\Delta = -0.93°$), $t_{2/3}(57) = 2.54$, $p = .007$, $d = 0.98$. The

Table 1 Overview of the instructions and results of the experiments. A short side-by-side summary of the instructions that were used (upper half), and the results that were obtained (lower half), separate for each experiment (columns)

	Experiment 1 Violation instructions	Experiment 2 Task-switching instructions	Experiment 3 Inversion instructions
Instructions			
Frequent task (75 %)	"Follow the rule"	"Task 1"	"Follow the rule"
Infrequent task (25 %)	"Break the rule"	"Task 2"	"Invert the rule"
Emphasis on rule violation	Yes	No	No
Number of instructed task sets	1	2	1
Results			
Immediate effect: response initiation	Strong	Weak	Strong
Immediate effect: response execution	Strong	Weak	Weak
Adaptation effect: response initiation	Yes	Yes	Yes
Adaptation effect: response execution	No	Yes	No

contrast between Experiment 1 and Experiment 3 was not significant, $t_{1/3} < 1$.

Movement times

Regarding the immediate effects, the comparison of the effect size of Experiment 1 ($\Delta = 43$ ms) against Experiment 2 ($\Delta = 10$ ms) was significant, $t_{1/2}(57) = 3.22$, $p = .001$, $d = 1.08$, as was the contrast between Experiment 2 and Experiment 3 ($\Delta = 34$ ms), $t_{2/3}(57) = 2.32$, $p = .012$, $d = 0.92$. The comparison between Experiments 1 and 3 was not significant, $t_{1/3}(57) < 1$.

For sequential effects, the comparison of Experiment 1 ($\Delta = -22$ ms) against Experiment 2 ($\Delta = 20$ ms) was significant, $t_{1/2}(57) = 2.90$, $p = .003$, $d = 0.91$, as was the contrast between Experiment 2 and Experiment 3 ($\Delta = -11$ ms), $t_{2/3}(57) = 2.17$, $p = .017$, $d = 0.64$. The contrast between Experiment 1 and Experiment 3 was not significant, $t_{1/3} < 1$.

Areas under the curve

Regarding the immediate effects, the comparison of the effect size of Experiment 1 ($\Delta = 14582\text{px}^2$) against Experiment 2 ($\Delta = 5200\text{px}^2$) was significant, $t_{1/2}(57) = 3.24$, $p = .001$, $d = 0.94$, as was the contrast between Experiment 1 and Experiment 3 ($\Delta = 9071\text{px}^2$), $t_{1/3}(57) = 2.90$, $p = .003$, $d = 1.09$. The comparison between Experiments 2 and 3 was not marginally significant, $t_{2/3}(57) = 1.33$, $p = .094$, $d = 0.55$.

For sequential effects, the comparison of Experiment 1 ($\Delta = 342\text{px}^2$) against Experiment 2 ($\Delta = 10789\text{px}^2$) was significant, $t_{1/2}(57) = 2.45$, $p = .009$, $d = 0.70$, as was the contrast between Experiment 2 and Experiment 3 ($\Delta = 3273\text{px}^2$), $t_{2/3}(57) = 1.78$, $p = .042$, $d = 0.60$. The contrast between Experiment 1 and Experiment 3 was not significant, $t_{1/3} < 1$.

Discussion

The direct comparison of the experiments allowed us to scrutinize the impact of the rule violation instructions as compared to both control conditions. Significant interactions between current response type and the between-subjects factor experiment showed that the impact of violations (Experiment 1) is more detrimental than the impact of task switches or inversions (Experiments 2 and 3). These differences, especially those between Experiment 1 and Experiment 3, can be solely attributed to the labeling of the actions as rule violations in the former experiment but not in the latter. It therefore seems as if simply relabeling the deviant response as an inversion instead of a violation is an effective way to minimize the impact of deviant responses.

There was also an apparent difference in the adaptation based on the previous response type between the experimental groups. While participants in the task-switching group were able to take parameters of the previous trial into account to adjust their performance on the current trial, this was only partly the case for both the violation group and the inversion group. Participants of those groups could adapt their response selection according to recent events, but failed to do so when it came to planning and executing the corresponding response. Here, the second violation or inversion in a sequence was just as slow and contorted, if not more, than the first one.

General discussion

In the present set of experiments, we investigated the impact that rule violations pose on the acting agent even when the rule in question is a simple S-R rule that is instantiated by instruction. We employed a two-dimensional finger-tracking task in which participants had to drag their finger from a starting area to one of two target areas on a touchscreen according to a pre-specified rule.

Summary of the results

In Experiment 1, we probed for the behavioral signature of rule violations regarding temporal and spatial parameters of the executed responses. In addition to analyzing how current rule violations influence participants' behavior, we also took previous experience with rule violations into account. We found a profound impact of current rule violations in both temporal and spatial measures. Rule violations took longer to be initiated and executed, and their movement trajectories were heavily bent towards the opposite side, which could indicate an ongoing influence of the original rule. And even though repeated rule violations were initiated with greater ease, we did not find any modulating influence of preceding rule compliance for measures capitalizing on response execution: repeated rule violations were as strongly affected by the original mapping rule as singular events of a rule violation.

In Experiment 2, we isolated the effect of rule violations by means of a first control condition by creating a task-switching experiment that called for the exact same behavior as Experiment 1. To this end, we instructed participants to respond in a frequent "Task 1" in most trials but prompted them to respond in an infrequent "Task 2" that was the reverse of the frequent task. Again, we found a strong temporal and spatial impact of the infrequent task set, but this time we also observed a profound sequential modulation: for repeated reversed responses, movement trajectories were as efficient as for responses based on the frequent task set.

Finally, in Experiment 3, we tested for inversions of an instructed task set and found similar sequential effects for the inverted responses compared to standard responses as we did for violations, whereas the overall impact of inversions was less pronounced than the impact of violations.

Comparison of the instructions

Even though at first sight, all three experiments employed the same task and used the same method, simple variations in the instruction caused strong differences in the participants' behavior. More precisely, the task that participants had to perform (c.f., Table 1) required finger movements to a left or a right target area based on the identity of a target stimulus. In some trials, participants had to aim for the exact opposite, and these infrequent trials were instructed either as "rule violation" (Exp. 1), "performing another task" (Exp. 2), or "inverting the rule". Even though the implied response was the same in all experiments, violating a rule turned out to be more difficult than inverting a rule or performing a reversed, separate task: rule violations were even slower in their initiation and execution and they were spatially more affected than both control conditions.

Before drawing conclusions about the possible mechanisms underlying rule violation, we would like to give a structured comparison of the instructions used (cf. Table 1). For one, the instructions differed as to whether one or two task sets were instructed, with Experiment 1 and 3 featuring only one task set and Experiment 2 featuring two distinct task sets. The main difference between the instructions involving one and the instructions involving two task sets is that, while two separate task sets allow for adaptation to the infrequent task, instructing only one task set seems to hinder participants from adjusting their performance based on recent events. The task sets for violations (Exp. 1) and inversions (Exp. 3) do not seem to be represented independently, but dependent on the frequent instruction. This representation *with strings attached* might cause the sequential modulation that we obtained here, which will be explained in more detail in the following sections.

For another, the instructions of Experiment 1 and Experiment 3 differed as to whether we emphasized that the infrequent task was not in accordance with the rule of the frequent task. For the violation instructions (Exp. 1), we specifically highlighted that the violation behavior ran counter to the original rule, whereas we did not use such an emphasis for the inversion instructions (Exp. 3). We will come back to this distinction in the following discussion.

Rule violations and cognitive conflict

The pronounced effects for rule violations as compared to rule-based responses accord with the idea that participants experience ongoing cognitive conflict during rule violations. Assuming that rules trigger automatic compliance (Asch 1956; Milgram, 1963), rule violations inherently provoke the activation of two response alternatives: the rule-based, automatic response and the planned violation response (for a recent perspective, see Kim & Hommel, 2015). The solution of this response conflict takes time, which explains the prolonged response initiation of violation responses. The conflict is not resolved completely, however, because the automatic, rule-based response still shapes violation behavior. This accounts for the ironic effect that rule violations are heavily influenced by the rule that participants try to violate (Wegner, 2009): they are confronted with a rule, however arbitrary, and they activate the corresponding response. Violations inherently include the recollection of the rule that has to be violated, so the activation of the rule-based response is strong enough that it cannot be suppressed entirely (Pfister et al., submitted). Thereby, we were able to isolate cognitive mechanisms that process (non-) conformity even in non-social settings.

Rule violations as a derived task set

What further differentiates violation behavior is how parameters of previous responses are taken into account. Alternatively, or in addition to the notion of cognitive conflict, the two response alternatives might be seen not as instances of the same task, but rather as distinct task sets. In this view, the observed adaptation after responses based on the infrequent task set might be taken to indicate task-switching effects (Allport, Styles, & Hsieh, 1994; Monsell, 2003; Rogers & Monsell, 1995). When simply switching to the infrequent task set, further responses based on this task set are easy, fast and efficient; parameters of previous responses are used to speed up the current response. But when violating rules, these parameters are not taken into account; a series of repeated rule violations poses repeated difficulty on the agent. This striking pattern of results was also found when participants were asked to invert a rule. If the current task set has to be derived from an instructed one and is not based on a separate, instructed task set (which is true for both, violations and inversions), this derived task set seems to be either (1) short lived and decays immediately, or (2) is not instituted as strongly as if it were an instructed task set or, moreover, (3) it could be used and attenuated immediately after finishing response execution, indicating repeated effort for repeated derivations of the currently relevant task set. Consequently, committing a violation response would always entail an immediate, endogenous switch back to rule-based responses (Arrington, & Logan, 2004; Arrington, Weaver, & Pauker, 2010; Kessler, Shencar, & Meiran, 2009; Liefooghe, Demanet, & Vandierendonck, 2010; Vandierendonck, Demanet, Liefooghe, & Verbruggen, 2012), as the derived task set for violations (and inversions) might not be as easily accessible or maintainable as an instructed task set.

A two-step activation model

This notion seems to be supported by research on how negations (and inversions) are represented in the cognitive system. Indeed, negations are assumed to be represented and retrieved in two separate steps: the non-negated concept is retrieved at first, followed by applying the negation for each individual retrieval process (Clark & Chase, 1972, 1974; Gilbert, 1991; Strack & Deutsch, 2004; Wegner, Coulton, & Wenzlaff, 1985). This holds true especially for negations that do not have a graspable meaning on their own, whereas negations seem to have only limited impact if participants can form an alternative representation (Hasson, Simmons, & Todorov, 2005; Fillenbaum, 1966; Mayo, Schul, & Burnstein, 2004). Even though such an alternative representation could have been formed in the present experiments (akin to the two task sets that we instructed in Experiment 2), the violation label might have worked against this tendency. Though speculative at present, this hypothesis seems to be a viable candidate for future research.

In any case, as violations and inversions produce the same sequential modulations, we propose that it is safe to say that the inversion of a rule (or in a broader picture: derivation, manipulation, negation, reformulation or modification of an existing task set) is one of the cognitive mechanisms that drive the behavioral parameters of rule violations. While this process partly explains the effects of rule violations, it does not drive them exclusively. Even though the sequential adaptation does not differ between these conditions, the burdens that violations pose on the agent at the moment of response execution exceed those of inversions.

Therefore, we conclude that in addition to this “cold cognition” explanation, it could be that “violate the rule” instructions have an emotional component (“hot cognition”), and participants might exhibit an active tendency to steer away from mental representations reflecting (socially) unwanted behavior. In this view, rule violations might be best described as an *inversion of an existing rule with an add-on*. Which components this add-on includes is an open question.

Conclusion

At any rate, the present set of experiments uncovers a striking pattern of challenges that rule violations pose on the acting agent. When regarded not as single instances, but as occurrences in a series of events, they are distinctly different from rule-based responses, not only in their behavioral markers, but also in the adaptation processes they trigger.

Of course, adaptations to recent instances of rule violations are only a first step toward understanding rule violations in context, and other questions wait to be addressed. How far do short-term memory effects influence the behavioral signature of rule violations (Hommel, Müsseler, Aschersleben, & Prinz, 2001)? Are agents able to adapt to violations if such behavior occurs over extended periods of time (akin to proportion congruency manipulations in studies on conflict adaptation; Logan & Zbrodoff, 1979; Notebaert, Houtman, Van Opstal, Gevers, Fias, & Verguts, 2009)? Can human agents profit from social support, e.g., by experiencing a co-actor to not stick to the rules (Milgram, 1963)?

Whatever the outcome of such investigations might be, our results already allow for a definite answer to the question raised in the introduction: are violations of simple S-R rules accompanied by a distinct behavioral signature? Even when all social aspects of the situation are removed, rule violations carry an observable behavioral signature.

As for now, we can only speculate how far the present results relate to real life situations involving non-conformity. “The veneer of civilization is very thin”, was what the curator of Marina Abramović’s Rhythm 0 commented. “What is absolutely terrifying is how quickly a group of people will become bestial if you give them permission to do so” (Chermayeff, Dupre & Matthew Akers, 2012). But if any rules apply—that is, if there is no permission for abnormal actions—a preference to obey these rules seems to rest tenaciously in us. Humans inherently experience difficulty when violating rules, even if removed from social influences, and this mode of operation seems to be shielded from becoming our default. This automatic adherence to rules might after all be what allows us to maintain social structures, and what the thin veneer of civilization is made out of.

References

Allport, D. A., Styles, E. A., & Hsieh, S. (1994). Shifting intentional set: exploring the dynamic control of tasks. In C. Umilta & M. Moscovitch (Eds.), *Conscious and nonconscious information processing: attention and performance XV* (pp. 421–452). Cambridge: MIT Press.

Arrington, C. M., & Logan, G. D. (2004). The cost of a voluntary task switch. *Psychological Science, 15*(9), 610–615.

Arrington, C. M., Weaver, S. M., & Pauker, R. L. (2010). Stimulus-based priming of task choice during voluntary task switching. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 36*(4), 1060–1067.

Asch, S. E. (1956). Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological Monographs: General and Applied, 70*, 1–70.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review, 108*(3), 624–652.

Chermayeff, M., Dupre, J., & Matthew Akers, M. (2012). *Marina Abramovic: the artist is present [motion picture]*. USA: Show Of Force.

Clark, H. H., & Chase, W. G. (1972). On the process of comparing sentences against pictures. *Cognitive Psychology, 3*(3), 472–517.

Clark, H. H., & Chase, W. G. (1974). Perceptual coding strategies in the formation and verification of descriptions. *Memory and Cognition, 2*(1), 101–111.

Fillenbaum, S. (1966). Memory for gist: some relevant variables. *Language and Speech, 9*(4), 217–227.

Gilbert, D. T. (1991). How mental systems believe. *American Psychologist, 46*(2), 107–119.

Gratton, G., Coles, M. G., & Donchin, E. (1992). Optimizing the use of information: strategic control of activation of responses. *Journal of Experimental Psychology: General, 121*(4), 480–506.

Hasson, U., Simmons, J. P., & Todorov, A. (2005). Believe it or not. On the possibility of suspending belief. *Psychological Science, 16*(7), 566–571.

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The theory of event coding: a framework for perception and action. *Behavioral and Brain Sciences, 24*, 849–878.

Kessler, Y., Shencar, Y., & Meiran, N. (2009). Choosing to switch: spontaneous task switching despite associated behavioral costs. *Acta Psychologica, 131*(2), 120–128.

Kim, D., & Hommel, B. (2015). An event-based account of conformity. *Psychological Science, 26*(4), 484–489.

Liefooghe, B., Demanet, J., & Vandierendonck, A. (2010). Persisting activation in voluntary task switching: it all depends on the instructions. *Psychonomic Bulletin and Review, 17*(3), 381–386.

Logan, G. D., & Zbrodoff, N. J. (1979). When it helps to be misled: facilitative effects of increasing the frequency of conflicting stimuli in a stroop-like task. *Memory and Cognition, 7*(3), 166–174.

Mayo, R., Schul, Y., & Burnstein, E. (2004). “I am not guilty” vs. “I am innocent”: successful negation may depend on the schema used for its encoding. *Journal of Experimental Social Psychology, 40*(4), 433–449.

Milgram, S. (1963). Behavioral study of obedience. *Journal of Abnormal and Social Psychology, 67*, 371–378.

Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences, 7* (3), 134–140.

Notebaert, W., Houtman, F., Opstal, F. V., Gevers, W., Fias, W., & Verguts, T. (2009). Post-error slowing: an orienting account. *Cognition, 111*(2), 275–279.

Pfister, R. (2013). *Breaking the rules: cognitive conflict during deliberate rule violations*. Berlin: Logos.

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., Janczyk, M., Wirth, R., Dignath, D., & Kunde, W. (2014). Thinking with portals: revisiting kinematic cues to intention. *Cognition, 133*(2), 464–473.

Pfister, R., Wirth, R., Schwarz, K., Steinhauser, M., & Kunde, W. (submitted). Burdens of non-conformity: motor execution reveals cognitive conflict during deliberate rule violations. *Cognition*.

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

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

Rogers, R. D., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General, 124*(2), 207–231.

Schroder, H. S., Moran, T. P., Moser, J. S., & Altmann, E. M. (2012). When the rules are reversed: action-monitoring consequences of reversing stimulus–response mappings. *Cognitive, Affective, and Behavioral Neuroscience, 12*(4), 629–643.

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review, 8*(3), 220–247.

Vandierendonck, A., Demanet, J., Liefooghe, B., & Verbruggen, F. (2012). A chain-retrieval model for voluntary task switching. *Cognitive Psychology, 65*(2), 241–283.

Wason, P. C. (1959). The processing of positive and negative information. *Quarterly Journal of Experimental Psychology, 11*(2), 92–107.

Wegner, D. M. (2009). How to think, say, or do precisely the worst thing for any occasion. *Science, 325*(5936), 48–50.

Wegner, D. M., Coulton, G. F., & Wenzlaff, R. (1985). The transparency of denial: briefing in the debriefing paradigm. *Journal of Personality and Social Psychology, 49*(2), 338.

Wirth, R., et al. (2015). Through the portal: Effect anticipation in the central bottleneck. *Acta Psychologica*. doi:10.1016/j.actpsy.2015.07.007.

Wirth, R., Pfister, R., & Kunde, W. (2015). Asymmetric transfer effects between cognitive and affective task disturbances. Cognition and Emotion, 1–18. doi:10.1080/02699931.2015.1009002.