Consciousness and Cognition 130 (2025) 103851

Contents lists available at ScienceDirect

Consciousness and Cognition

journal homepage: www.elsevier.com/locate/yccog

Full Length Article

Temporal binding during deliberate rule breaking

Carl Michael Galang[a,*], Ayça Akan[a], Roland Pfister[b], Marcel Brass[a,*]

[a] Social Intelligence Lab, Department of Psychology, The Berlin School of Mind and Brain, Humboldt-Universität zu Berlin, Germany
[b] INTACT Lab, Department of Psychology, Trier University, Germany

A R T I C L E I N F O

Keywords:
Temporal binding
Intentional binding
Rule breaking
Sense of agency
Multisensory integration
Interval estimates

A B S T R A C T

Rules are deeply ingrained in our cognition. The current study investigates the influence of rule breaking on explicit sense of agency as well as the implicit perceptual illusion of temporal binding. Participants completed a free choice task that involved following or breaking a predetermined rule. The task required pressing a key that matched to a visual stimulus which triggered a corresponding change after a delay. Participants estimated the delay as an index of temporal binding. The results showed similar levels of explicit agency for rule following and breaking. Temporal binding, by contrast, was indeed influenced by rule breaking; however, the relationship is complex. Specifically, participants had smaller interval estimates for rule following vs. breaking at the 100 ms delay, likely reflecting cognitive conflict during rule breaking, whereas this effect reversed for the 400 ms and 700 ms delays. We interpret our results in relation to the wider rule breaking and temporal binding literature.

1. Introduction

From a young age we are taught rules that govern our behaviours. Recent work has shown that such rule following behaviour is so ingrained in our cognitive architecture that choosing to break a rule can incur a measurable cognitive cost. This is even the case if rules are arbitrary and rule breaking is not penalized (Pfister et al., 2016). For instance, participants were instructed to move their mouse cursor to a target on the upper-left hand side of the screen in response to one symbol, but if they saw a different symbol, they should instead move to the upper-right hand side of the screen. Crucially, participants could choose whether they wanted to follow or break this rule on every trial. Breaking the rule entailed performing the wrong action for a given symbol. Cursor movements (measured via the trajectory of the mouse from the start and final positions) during the rule breaking trials were biased towards the correct final position. In other words, it was as if the participants' actions were implicitly being pulled towards the rule following behaviour even when they intentionally broke the rule. This effect was even more striking when compared to another set of participants that were told to respond according to a reversed rule rather than to break it – that is, participants completed the same task with the only change being the framing of the task ("apply an alternative rule" vs. "break the rule"). With the reversed rule framing in mind, participants no longer showed a notable movement bias. Other effects were also reported such as participants exhibiting a significant preference for following the rule vs. breaking it, along with quicker task initiation and faster movements during rule following, which all suggests that there is a strong bias towards following rules.

The operationalization chosen in Pfister et al. (2016) reduces rule-breaking to intentionally performing the opposite of what a rule prescribes (Pfister, 2022). This operationalization comes with the elegant property that for every single action it is clear whether a

* Corresponding authors.
E-mail addresses: galang.michael@gmail.com (C.M. Galang), marcel.brass@hu-berlin.de (M. Brass).

https://doi.org/10.1016/j.concog.2025.103851
Received 28 October 2024; Received in revised form 1 February 2025; Accepted 17 March 2025
Available online 21 March 2025
1053-8100/© 2025 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

participant intentionally violated a mapping rule or whether they committed an inadvertent error. At the same time, this operationalization can be questioned in that rule breaking is embedded in a *meta*-rule that asked participants to choose between following or violating a rule (for a corresponding critique, see Gozli, 2017, 2019). The prompt that solicited rule violation behavior thus came from the same source as the original mapping rule, because both were issued by the experimenter. Moreover, rule violation came with a necessary negation operation so that issues during rule violation might simply reflect negation processing rather than actually reflecting peculiarities of rule violation behavior. Both issues were addressed and resolved in follow-up work.

For instance, similar signatures of cognitive conflict during rule breaking emerged when participants performed unsolicited rule violations to expedite task completion (Pfister et al., 2019). Participants of this study completed a virtual maze game in which they had to deliver virtual pizza by steering a bicycle courier through a 2D city map. These maps offered the possibility of using forbidden shortcuts by cycling through a public park rather than taking a detour on normal roads. Navigation was performed with the arrow keys on the computer keyboard, and inter-keystroke intervals increased when participants purposely used forbidden shortcuts, indicative of cognitive conflict when initiating a rule violation. The role of negation processing was further targeted in a study that used finger-tracking on iPads to record movement trajectory data (Wirth et al., 2016). This study fully capitalized on the strategy of using a *meta*-rule to solicit rule violations by asking participants on every trial to either follow or break the rule (for a similar procedure, see Exp. 2 in Pfister et al., 2016). Crucially, in another condition, participants performed the very same actions but were instructed to follow the rule or reverse the rule (Exp. 3 in Wirth et al., 2016). Attraction of the movement trajectories during rule violation was substantially larger than for reversed (negated) rules, thus indicating that the simple fact of labelling an action as a rule violation has profound consequences for how this action is processed. Converging evidence for this conclusion comes from a priming task which showed that the simple label of performing a rule violation (even one that is fully solicited by the experimenter) leads to increased sensitivity towards negatively valanced and authority-related probes words (Wirth et al., 2018a). Rule breaking, therefore, not only comes with cognitive conflict due to a continued representation of the original rule, but it also seems to imply hyperactive monitoring of one's performance, as well as implicit expectations of punishment after breaking a rule.

An interesting and unexplored aspect of rule breaking is its relationship with the sense of agency. The sense of agency is the feeling of control we have over our actions and subsequent outcomes (Moore & Obhi, 2012; Haggard, 2017). Given that the previously reviewed work suggests that rule breaking influences actions, it seems likely that one's sense of agency will also be affected by deliberate rule breaking. However, the direction of the effect is not so obvious – on one hand, one could make a case that deliberately breaking a rule requires effort and should thus lead to strong sense of agency, as there is some evidence to suggest that exerting physical effort boosts the sense of agency (Demanet et al., 2013). On the other hand, rule breaking leads to cognitive conflict (Pfister et al., 2016, 2019; see also Wirth et al., 2018b). This could in turn mean that deliberate rule breaking leads to a lower sense of agency, as there is some evidence to suggest that action selection disfluency (which results from cognitive conflict) can attenuate agency (Chambon & Haggard, 2012; Vastano et al., 2017).

Sense of agency can be measured explicitly by having participants rate their subjective level of control over their actions and corresponding changes in the environment. Additionally, previous research has often targeted the perceptual illusion of temporal binding (originally referred to as intentional binding; Haggard et al., 2002). Temporal binding refers to perceiving one's actions and the consequences of those actions as attracted to one another compared to when such actions are not intentional or do not occur at all (e.g., Haggard et al., 2002). While early work has generally accepted temporal binding as a measure of *implicit* agency (Moore & Obhi, 2012; Haggard, 2017; Malik et al., 2022 for reviews), this interpretation is not without controversy (e.g., Suzuki et al., 2019; Kirsch et al., 2019); we discuss this point more in the General Discussion. To investigate the effects of deliberate rule breaking on explicit agency and temporal binding, we had participants complete a free choice task where they could either follow or break a predetermined rule. The task itself (if following the rule) involved pressing one key if they saw a square or another key if they saw a triangle; depending on their choice, either a left or right circle would then change colour on the screen. Crucially, to get a measure of temporal binding, there was a delay between their keypress and the colour change. At the end of each trial, participants would then be asked to estimate this delay (between 1–1000 ms) as an index of temporal binding. Given the previously discussed works, we cannot make any directional predictions regarding the effects of rule breaking on temporal binding – according to the "effort hypothesis", rule breaking should lead to smaller interval estimates compared to rule following; however, according to the "cognitive conflict hypothesis", we should see the opposite. Moreover, we had participants provide self-reported levels of control as a measure of *explicit* sense of agency during the task.

Experiment 1 was our first foray exploring this topic; however, as discussed in the Transparency and Openness section below, unforeseen problems arose that forced us to significantly deviate from our pre-registration. Experiment 2, then, was used to corroborate the results of Experiment 1. We also performed secondary analyses on reaction times and rule choice – this was done to validate our methods, as previous work (e.g., Pfister et al., 2016) suggests that participants should not only heavily favour rule following over breaking, they should also be faster to respond for rule following.

2. Experiment 1

2.1. Methods

2.1.1. Transparency and Openness

The raw and processed data, processing scripts, analyses, and pre-registration can be found on our OSF repository: https://osf.io/fnmc2/. Please note that we deviated from our original analysis plan due to unforeseen issues regarding participant retention rates. In particular, we did not expect that a large number of participants would only choose to follow the rule and never break it. As per our pre-registered exclusion criteria, participants with fewer than 5 trials in any condition (Rule x Time Delay) would be excluded from the study. As a result, a large number of participants were removed. Indeed, we kept collecting participants to keep replacing those excluded; however, we capped ourselves to 400 due to resource constraints. Out of the 400 total, in the end, only 167 were valid. We also pre-registered a 2x3 RM ANOVA for our primary analysis (with Rule – Break, Follow, and Delay – 100 ms, 400 ms, 700 ms, as the factors). However, given the extreme loss of participants and the uneven trial numbers between rule break and follow conditions, we realized this would not be an appropriate approach to our data (as a traditional RM ANOVA works on averaged data per condition per participant, it cannot account for the uneven trial numbers per condition nor allow the inclusion of missing data for a condition). The solution we opted to tackle these issues with was Linear Mixed-Effect Models (LMM). LMM's have the advantage of being able to analyze trial-level data and can handle uneven designs, including participants with missing data for a given condition (e.g., Magezi, 2015; Schielzeth et al., 2020; Brown, 2021). Given this, we also removed our rule regarding 5 trials in order to use more of our data (going from 167 to 273 valid participants). As these changes represent a major deviation to our original pre-registration, we followed up our first experiment with a direct replication and pre-registered the new methods described.

2.1.2. Participants

Our original plan was to collect n = 200, which was based on an a priori power analysis conducted via G*Power (Faul et al., 2007; 2009) assuming dz = 0.2 at 80 % power for a pairwise *t*-test (note that our original prediction was for a main effect of Rule Breaking vs. Following, with the factor Delays added due to convention in the temporal binding literature and, initially, for exploratory purposes). A dz of 0.2 was chosen following Cohen's guidelines for a small effect size (note that the main effect in this analysis is mathematically identical to a paired *t*-test so that we based our power calculations on a paired *t*-test; Langenberg et al., 2023). In the end, due to the issues described above (see Transparency and Openness section), 400 participants were collected via Prolific; and, after applying our pre-registered exclusion criteria (minus the less than 5 trials per condition criteria, as discussed above), 273 participants (Male = 146, Female = 120, Other = 7; Mean Age = 30.8, SD Age = 9.4) were left for final analysis. All participants were recruited via Prolific (prolific.co), as there is evidence to suggest that samples obtained from this platform provide more reliable responses compared to MTurk and other alternative platforms (e.g., Peer et al., 2017; Palan and Schitter, 2018). Participants were filtered for fluency in English and compensated £4 for participation (30 min study, equating to an hourly rate of £8).

2.1.3. Apparatus, Stimuli, & procedure

The experiment was programmed in JavaScript using the JsPsych library (De Leeuw, 2014) and administered online using JATOS (Lange et al., 2015). Participants were redirected to the JATOS hosted experiment website via Prolific. At the start of the experiment, participants were given the chance to read over the letter of information (which contained information about the study, the researchers, and their rights as participants) and give consent. They were then prompted to answer a series of demographic questions related to their age, gender, and ethnicity, as well as provide their prolific ID. Participants were then shown the instructions page, which indicated that on each trial they would see one of two geometric symbols: a triangle or a square. Each symbol would be assigned to a keypress as a rule – depending on the geometric symbol shown, they were instructed to press either the "A" or "D" key on their keyboard with their left-hand. Afterwards, they were shown an example of a single trial: they would see the square or triangle in the middle of the screen with two white circles with black outlines to the upper left or right of the square/triangle. Regardless of their choice to break or follow the rule, the "A" key would be associated with the upper left circle and the "D" key to the upper right circle, and that pressing the key would make the associated circle change colour from white to blue after some delay (see Fig. 1).

A single trial would consist first of a prompt for participants to choose whether they would break or follow the rule. Afterwards, they would see either a square or a triangle with the two circles above. Participants would also be reminded of their choice at the bottom of the screen (e.g., "Your choice: Follow Rule"). Participants would then press "A" or "D", depending on the symbol and their choice to follow or break the rule. After some delay, the appropriate circle would turn blue. Participants were told that the delay could be anytime between 1–1000 ms; however, in reality there were only three delays: 100 ms, 400 ms, and 700 ms (fully randomized).

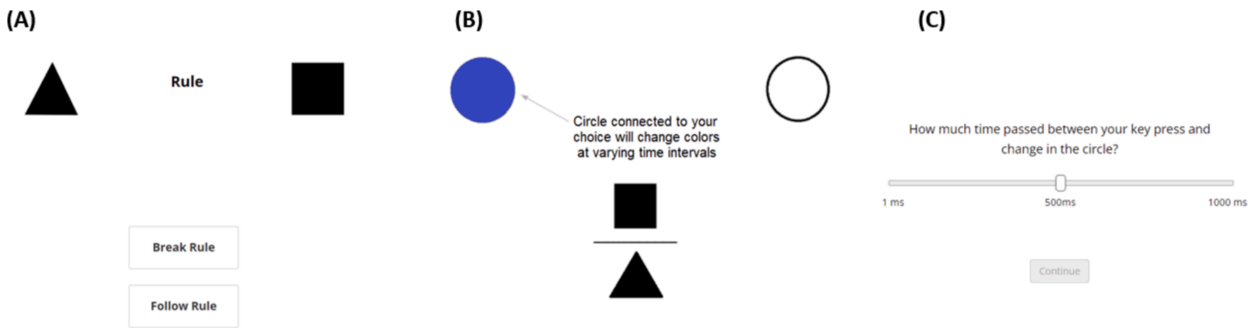

Fig. 1. Instruction screens for an exemplary trial. (A) Participants were prompted to choose to break or follow a rule by clicking the appropriate button on the screen. (B) Participants were randomly shown one of the two geometric symbols (a square or triangle), and their keypress (depending on whether they followed or broke the rule) changed one of the two circles on the screen. (C) At the end of the trial they were asked to rate "How much time passed between your keypress and change in the circle?" on a visual analogue scale ranging from 1 ms to 1000 ms.

Afterwards, they were shown a visual analogue scale ranging from 1 ms to 1000 ms with the prompt: “How much time passed between your key press and the change in the circle?”. Participants had to move the slider at least once before the continue button was enabled (to prevent participants from simply clicking “continue” without engaging with the task).

Participants first started with 6 practice trials to familiarize themselves with the task. After each trial, they were shown whether they pressed the key that corresponded to their choice to break or follow the rule to provide feedback. To make sure participants believed that the delay could be between 1–1000 ms, for the practice trials; we did randomize the delay as such. Participants would then be given feedback to how close their estimate was after each trial. Afterwards, participants were informed that they would start the main part of the experiment and that no feedback would be given. The main part of the experiment consisted of three blocks of 30 trials. Every 5 trials, participants would also be prompted to rate how much control they felt for the colour change with a visual analogue scale ranging from “no control” to “a lot of control” (for the trial they just completed). This was added as a secondary measure of explicit judgments of agency. At the end of the experiment, participants were asked some questions regarding how guilty they feel when breaking rules, how often they violate rules in general, and how much they disapproved of others for breaking rules (answered via a visual analogue scale ranging from 0 to 100). These latter questions were added primarily for exploratory reasons, and we do not present them in the current manuscript but include them in a supplementary materials file.

2.1.4. Design and analysis plan

As discussed in the Transparency and Openness section above, while we initially pre-registered the use of a traditional RM ANOVA (with Rule – Break, Follow, and Delay – 100 ms, 400 ms, 700 ms as the independent variables) for our main analysis, the unexpected difficulties with participant retention along with the major imbalance between the Break and Follow conditions made us consider other options. In the end, we opted to conduct a Linear Mixed-Effects Model to analyze our results. Final statistical analyses were conducted on R (ver. 4.4.1) using the lme4 (Bates et al., 2015) and emmeans (Lenth, 2024) packages. Visualization was conducted using the ggplot2 (Wickham, 2016) and sjPlot (Ludecke, 2024) packages. Alpha was set using the conventional $p < 0.05$ criterion. We fit a linear mixed-effects model via the lme4 package in R with the following structure:

Interval Estimates ~ Rule * Delay + (Rule + Delay | ID).

This model includes fixed effects for Rule, Delay, and their interaction, as well as random intercepts and slopes for rule and delay across participants (ID). The model used sum contrast coding, with Rule Breaking and Delay at 100 ms as the reference categories. As we are primarily interested in the overall effect of each factor (rather than the specific coefficients for each term in the model), we performed an omnibus test via the use of the anova function on our mixed-model in R. Follow-up pairwise comparisons were conducted via the emmeans package.

As per our pre-registration, we removed trials with incorrect responses (i.e., broke the rule when they indicated they would follow it) and where response times deviate more than 2.5 standard deviations from the cell mean (calculated per participant and condition). On average, these criteria resulted in 10.9 % of trials removed.

In addition to interval estimates, we also performed secondary analyses on reaction times, rule choice, and explicit judgments of agency. Reaction times and explicit judgments of agency were analyzed using the following lme4 models:

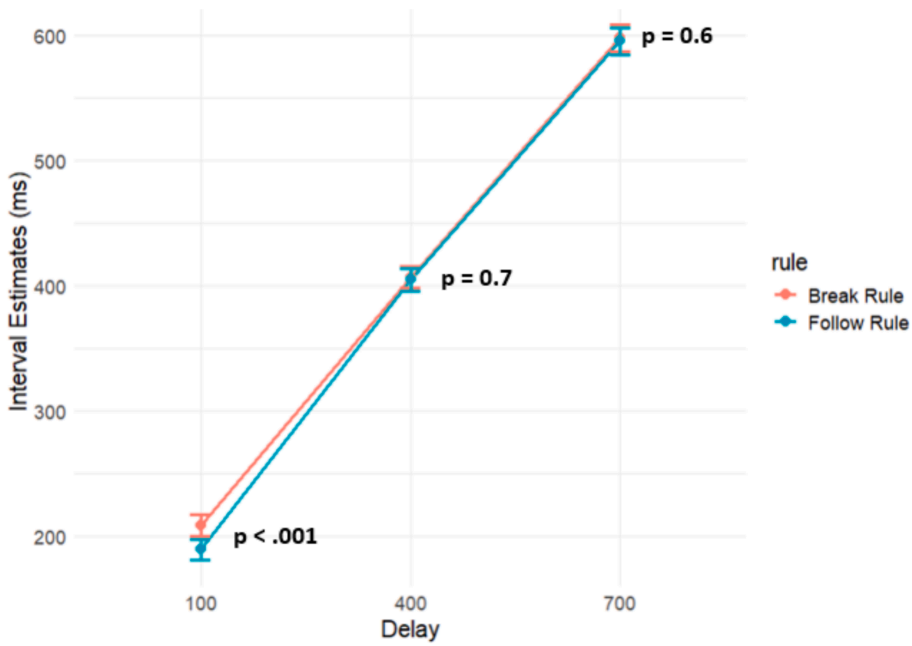

Fig. 2. Interval estimates in Experiment 1 as a function of experimental condition Error bars represent SEM. P-values have been corrected for multiple comparisons via the Bonferroni-Holm method.

Reaction Time ~ Rule + (Rule | ID).

Judgment Ratings ~ Rule*Delay + (Rule + Delay | ID).

Note that delay is not added in the model for reaction times, as the delay occurs only after the keypress. Finally, to investigate if there is bias in participant decision to break or follow a rule, the following lme4 model was used (using the glmer function for a logistic regression due to the binary nature of the outcome variable):

Rule ~ 1 + (1 | ID).

2.2. Results

2.2.1. Interval estimates

Fig. 2 shows interval estimates as a function of all experimental conditions. There was a significant main effect of Rule [$F(1, 197.2) = 4.92$, $p = 0.028$, $\eta_p^2 = 0.02$], wherein participants, overall, had smaller interval estimates when following the rule [M = 396 ms, SE = 7.8] vs. breaking it [M = 404 ms, SE = 7.4]; a statistically significant main effect of Delay [$F(2, 294.3) = 582.40$, $p < 0.001$], wherein participants, overall, showed the smallest interval estimates at the 100 ms delay [M = 199 ms, SE = 7.9], increasing at the 400 ms delay [M = 406 ms, SE = 8.4], and being the largest at the 700 ms delay [M = 596 ms, SE = 10.5]; and, finally, a statistically significant Rule x Delay interaction effect [$F(2, 19314.18) = 5.37$, $p = 0.004$]. To decompose this interaction, we performed post-hoc pairwise comparisons of Break vs. Follow for each Delay condition (as per our pre-registration, each p-value was corrected via the Bonferroni-Holm method). The results show that there is a statistically significant difference between Break and Follow at the 100 ms Delay [$t(742) = 3.86$, $p < 0.001$], wherein participants had smaller interval estimates when following the rule [M = 190, SE = 8.1] compared to breaking it [M = 209, SE = 8.5]. The difference between Break and Follow were not statistically significant at the 400 ms [$p = 0.7$] and 700 ms [$p = 0.6$] delay conditions.

2.2.2. Reaction times

The results showed a statistically significant main effect of Rule [$F(1, 205.5) = 200.4$, $p < 0.001$, $\eta_p^2 = 0.49$]. Participants were faster when following the rule [M = 658 ms, SE = 15.4] compared to when they were breaking it [M = 772 ms, SE = 16.4].

2.2.3. Explicit judgments of agency

The results do not show a statistically significant effect of Rule [$F(1, 197.1) = 2.76$, $p = 0.1$, $\eta_p^2 = 0.01$] nor the interaction effect [$F(2, 32101.8) = 0.75$, $p = 0.48$, $\eta_p^2 < 0.001$]. However, the results did show a statistically significant effect of Delay [$F(2, 320.3) = 26.6$, $p < 0.001$, $\eta_p^2 = 0.14$]. This effect shows that participants rated the highest level of control with the 100 ms delay [M = 49.2, SE = 1.6], which decreased with both the 400 ms [M = 43.7, SE = 1.5] and 700 ms delays [M = 42.4, SE = 1.5].

2.2.4. Rule choice

The results of the generalized linear mixed-effects model showed that the estimated log-odds of the outcome (Rule Following, as Rule Breaking was coded as 0 and is the reference) is 1.58 and is statistically significant [$z = 12.8$, $p < 0.001$]. This corresponds to an odds ratio of 4.87, indicating that participants are 4.87 times more likely to follow the rule than break it throughout the experiment.

2.3. Discussion

The results of Experiment 1 indicate no evidence that rule breaking affects subjective ratings of control (despite ratings being sensitive to delay in replication of previous findings; Ebert & Wegner, 2010; Galang, Cracco, Chirkov, Obhi, & Brass, 2024). However, rule breaking did indeed influence temporal binding. Specifically, at the 100 ms delay, participants showed stronger binding effects (i.e., smaller interval estimates) when they followed the rule vs. when they broke it. However, this difference disappeared with the later delays. This could indicate that, at least for very short delays between action and outcome, the cognitive conflict triggered by choosing to break the rule negatively influenced one's interval estimates. Of course, before continuing to interpret this effect, the fact that the methods and analysis had to be drastically changed in this experiment leaves room for doubt about the reliability of the results. Furthermore, while we predicted a main effect of Rule Breaking, we did not power for an interaction effect. As such, we conducted a second experiment to directly replicate this finding. Experiment 2 follows Experiment 1 closely; however, we made two modifications – first, we formally removed the minimum 5 trials per condition requirement in our new pre-registration, and second, we prompted participants to try and be even with their choice to follow and break the rules. The latter was done in an attempt to get more of a balanced data set between rule following and breaking conditions.

Our secondary analyses also yielded interesting results. The reaction times and rule choice data corroborate previous rule breaking work (Pfister et al., 2016), which predicts that participants should have both a strong bias towards rule following and be faster at it.

3. Experiment 2

3.1. Methods

3.1.1. Transparency and Openness

The raw and processed data, processing scripts, Jamovi analyses, and pre-registration can once again be found on OSF: https://osf.io/fnmc2/.

3.1.2. Participants

As per our pre-registration, we aimed to collect n = 250. We used the dataset from Experiment 1 to model power across multiple sample sizes (via the mixedpower package on R; Kumle et al., 2021). We specifically looked to power our study to find the Rule x Delay interaction effect seen in Experiment 1 for the measure of temporal binding (note that we did not consider other tests, and this is a limitation in our design – see limitations in the General Discussion). However (as explained in our pre-registration), we expected to lose participants based on our exclusion criteria. As such, we planned to replace lost participants up until a total maximum of 300 recruited participants – once 300 is reached, regardless of how many valid participants we have, we stopped data collection. After applying our pre-registered exclusion criteria – wherein any participant that did not show a statistically significant correlation between temporal estimates and interval delay, along with participants that had more than 20 % of their trials removed, would be excluded – we were left with n = 236 [Male = 120, Female = 114, Other = 2; Mean Age = 27.7, SD Age = 11] for final analysis. All participants were recruited via Prolific (prolific.co). Participants were filtered for fluency in English, geographically limited to the UK, and compensated £3.35 for participation. We opted to limit our participant pool to the UK for Experiment 2 in order to better control for potential differences in interpretations of rule breaking, and thus reduce variability and enhance the reliability of our findings. The absolute compensation was changed for experiment 2 to match the fact that participants completed the experiment 10 min faster than anticipated in Experiment 1 – this new time estimate of 20 min was then used for experiment 2; however, since £3.35 equates to ~£8.04 per hour for 20 min, the relative hourly rate does not change from Experiment 1 to 2.

3.1.3. Apparatus, Stimuli, & procedure

There was no change from Experiment 1, other than the already discussed prompt to try and be even with their choice to follow and break the rule.

3.1.4. Design and analysis plan

There was no change from Experiment 1 for our main analysis. However, for the secondary analysis involving explicit judgments of agency, the model involving both Rule and Delay as random slopes would not converge. As such, we were left with the following lme4 model:

Judgment Ratings ~ Rule*Delay + (1 | ID).

3.2. Results

3.2.1. Interval estimates

Fig. 3 shows interval estimates as a function of all experimental conditions. The results showed a statistically significant main effect of Rule [$F(1, 206.8) = 5.01$, $p = 0.02$, $\eta_p^2 = 0.02$]; in contrast to Experiment 1, participants had larger interval estimates when following the rule [M = 432 ms, SE = 6.1] vs. breaking it [M = 424 ms, SE = 6.4]; a statistically significant main effect of Delay [$F(2, 235.3) = 1382$, $p < 0.001$], wherein participants, overall, showed the smallest interval estimates at the 100 ms delay [M = 137 ms, SE = 7.4],

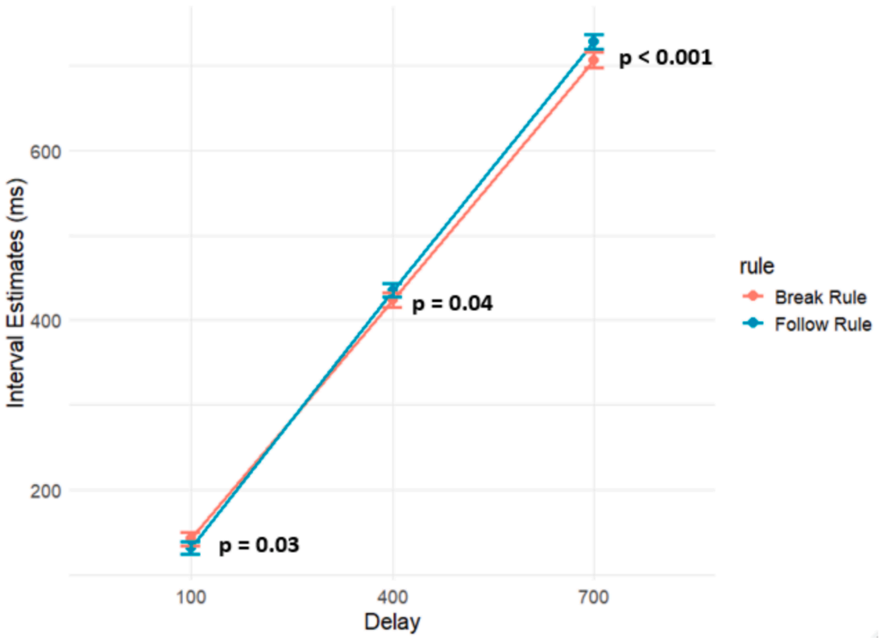

Fig. 3. Interval estimates in Experiment 2 as a function of experimental condition. Error bars represent SEM. P-values have been corrected for multiple comparisons via the Bonferroni-Holm method.

increasing at the 400 ms delay [M = 429 ms, SE = 7.9], and being the largest at the 700 ms delay [M = 718 ms, SE = 8.6]; and, finally, a statistically significant Rule x Delay interaction effect [F(2, 18558) = 14.12, $p < 0.001$]. To decompose this interaction, we performed post-hoc pairwise comparisons of Break vs. Follow for each Delay condition (as per our pre-registration, each p-value was corrected via the Bonferroni-Holm method). Following Experiment 1, the results showed that there is a statistically significant difference between Break and Follow at the 100 ms Delay [t(973) = 2.18, $p = 0.03$], wherein participants had smaller interval estimates when following the rule [M = 131, SE = 7.6] compared to breaking it [M = 142, SE = 7.9]. Unlike Experiment 1, however, we now find a reversal of this effect at 400 ms [t(910) = 2.36, $p = 0.04$] and 700 ms [t(902) = 4.46, $p < 0.001$] both showing that participants had larger interval estimates when following the rule [400 ms: M = 435 ms, SE = 8.1; 700 ms: M = 729 ms, SE = 8.8] compared to breaking it [400 ms: M = 423 ms, SE = 8.5; 700 ms: M = 707 ms, SE = 9.0].

3.2.2. Reaction times

The results showed a statistically significant main effect of Rule [F(1, 219.5) = 138, $p < 0.001$, $\eta_p^2 = 0.39$]. This shows that participants were faster when following the rule [M = 833 ms, SE = 27] compared to when they were breaking it [M = 966 ms, SE = 28].

3.2.3. Explicit judgments of agency

The results do not show a statistically significant effect of Rule [F(1, 3568.1) = 0.57, $p = 0.45$, $\eta_p^2 < 0.01$] nor the interaction effect [F(2, 3561.7) = 0.75, $p = 0.18$, $\eta_p^2 < 0.001$]. However, as with Experiment 1, the results did show a statistically significant effect of Delay [F(2, 3561.5) = 13.3, $p < 0.001$, $\eta_p^2 = 0.007$]. This effect shows that participants rated the highest level of control with the 100 ms delay [M = 40.9, SE = 1.9], which decreased with both the 400 ms [M = 39.1 SE = 1.9] and 700 ms delays [M = 38.4, SE = 1.9].

3.2.4. Rule choice

The results of the generalized linear mixed-effects model showed that the estimated log-odds of the outcome (Rule Following, as Rule Breaking was coded as 0 and is the reference) is 0.28 and is statistically significant [z = 6.8, $p < 0.001$]. This corresponds to an odds ratio of 1.3, indicating that participants are 1.3 times more likely to follow the rule than to break it throughout the experiment.

3.3. Discussion

The results of Experiment 2 provide both similarities and differences from Experiment 1. The result of the interval estimates at 100 ms directly replicate those found in Experiment 1 – participants seemed to show stronger temporal binding effects via estimating smaller intervals when following the rule vs. breaking it. However, whereas Experiment 1 did not find any evidence for a difference between following and breaking the rule at 400 ms and 700 ms delays, the results of Experiment 2 indicated that the effects actually reverse for these delays: participants had smaller interval estimates when they broke the rule rather than followed it. In the General Discussion, we go over possible interpretations for this reversal effect.

In regard to all remaining analyses, the results generally follow Experiment 1. Participants were faster to respond when following the rule vs breaking it. Participants once again did not show any difference in the explicit judgments of agency; however, they again showed a linearly decreasing effect as the delays became longer. Finally, although participants seemed to heed our instructions to try and make an even decision to follow or break the rule (as indicated by the smaller odds ratio in Experiment 2 compared to 1), there was still a statistically significant bias towards rule following.

4. General discussion

Our overall research aim was to investigate if and how explicit sense of agency and temporal binding, ostensibly an implicit measure of the sense of agency, are influenced by rule breaking. The results of both Experiment 1 and 2 indicate that we did not find enough evidence to suggest that explicit agency is influenced by rule breaking; rather, the results seem to suggest that an influence indeed exists on the implicit measure of temporal binding; however, the direction of this effect is surprisingly complex. The most robust finding across both experiments were the interval estimates at the 100 ms delay – participants in both experiments showed smaller interval estimates when they followed the rule vs. when they broke it. This could be explained by the fact that rule breaking has been shown to lead to cognitive conflict (e.g., Pfister et al., 2016; Wirth et al., 2018b) which may be accompanied by or causes action disfluency – this is evident in our secondary results across both experiments showing that participants had both a bias to choosing to follow the rule and had faster reaction times when doing so. As action disfluency has been shown to attenuate temporal binding (Vastano et al., 2017), it is possible that interval estimates during rule breaking were likewise influenced by this effect (i.e., became larger). As briefly mentioned in the introduction, however, there is no consensus that temporal binding measures the sense of agency (e.g., Schwarz et al. 2019). Alternative explanations suggests that such temporal binding effects could arise as the result of a general mechanism related to perceived cause-effect relationships in the world (e.g., Hoerl et al., 2020), as the result of multi-sensory integration (Klaffehn et al., 2021; Wolpe et al., 2013; but see Klaffehn et al., 2024), or due to procedural confounds in common designs to assess temporal binding (Gutzeit et al., 2023). Note that some recent work still continues to use temporal binding as a measure of implicit agency (e.g., Galang et al., 2021; Spaccasassi et al., 2023; Ciaunica et al., 2024; Mariano et al., 2024); and indeed, Wiesing & Zimmermann (2024) recently found results consistent with earlier work on temporal binding (i.e., stronger effects for intentional actions compared to non-intentional; see also Weller et al., 2020). We present this broader context as we think it crucial for the reader to understand the current controversy with its use as a measure of the sense of agency. And as discussed below, a non-agency interpretation of our results will also be presented.

In contrast to the 100 ms delay, the interval estimates at the 400 ms and 700 ms delays are harder to explain. Experiment 1 found no evidence of a statistically significant difference between following vs. breaking the rule in these delays, while Experiment 2 did. Furthermore, not only was a statistically significant difference found, but the effect also went in the other direction as compared to the 100 ms delay: participants had smaller interval estimates when they broke the rule vs. when they followed it. In isolation, one could explain this by pointing to the fact that choosing to break a rule is not the natural choice and thus requires extra effort to do so. And as effort (at least physical effort; see: Demanet et al., 2013) has been shown to lead to stronger binding effects, it is possible that the same thing occurred for rule breaking in Experiment 2. Of course, the difficulty with these results is that one needs to explain them in tangent with those found for the 100 ms delay.

As far as we are aware, while there are plenty of examples in the temporal binding literature of an effect exclusively occurring to a specific delay, no other temporal binding study has reported a *reversal* of effects. We speculate that this reversal effect could be explained by increased response caution – in general, participants showed slower reaction times in Experiment 2 compared to Experiment 1. This was probably due to the fact that participants in Experiment 2 were instructed to be more even with their choices, and indeed, participants in Experiment 1 were 4.87 times more likely to follow the rule vs. break it, while participants in Experiment 2 were only 1.3 times more likely to do the same. Given the extremely strong bias towards rule following, participants in Experiment 2 may have slowed down their response times overall in order to better track their frequency of rule breaking. As such, perhaps response caution allowed for the effects of effort (as discussed above) to be seen in Experiment 2 for the later delays. This account then suggests a two process model, one wherein cognitive conflict via rule breaking influences short delays at 100 ms, which then dissipates as the delays become longer and is subsequently replaced by an effort effect, thus reversing the direction of the interval estimates. Note that this interpretation of quickly diminishing conflict does not clash with observations that showed performing a rule violation to reduce conflict effects on following, unrelated interference tasks (Wirth et al., 2018b). Instead, work on conflict adaptation suggests that conflict indeed dissipates quickly while it upregulates sustained control efforts that affect performance in upcoming action episodes (Yeung et al., 2004). We make explicit here that this is a post-hoc explanation of an unexpected finding, and as such, future work will be needed to confirm this interpretation.

Another possible explanation is that these effects are due to more regression to the mean during rule breaking. Vierordt's Law suggests that regression to the mean is particularly pronounced in time perception, where interval estimates gravitate towards the middle interval (which can be seen in temporal binding studies, e.g., Zimmermann & Cicchini, 2020; Wiesing & Zimmermann, 2024). In the case of Experiment 2, the middle interval is 400 ms – as such, one could interpret the results as varying degrees with which participant judgements gravitate towards 400 ms (including estimates with 400 ms delay, as there will be some error). Looking at the results of Experiment 2, one can see that there is a dragging effect – estimates at the 100 ms delay are being dragged up, while estimates at the 400 ms and 700 ms delay are being dragged down. One explanation of the results then is that rule breaking leads to stronger dragging effects. This may be due to the fact that cognitive conflict and the extra motivation and attention it takes to choose to rule break leads to a higher chance of biases influencing reasoning (e.g., De Neys, 2006). The advantage of this explanation is that it does not need to posit multiple systems at play, as it explains the difference in rule breaking and following across all delays. However, it should be noted that this interpretation goes against some work that suggests that rule breaking leads to hyper-monitoring (Wirth et al., 2018a), which in turn should lead to more accurate interval estimates.

Finally, in regard to our measure of explicit sense of agency (i.e., control ratings), we did not find any evidence to suggest that this measure is influenced by rule breaking. However, we do not rule out the possibility that other experimental designs involving rule breaking may influence it. This is especially true for cases in enriched scenarios above and beyond simple keypress-response paradigms (e.g., Schwarz, Tonn, Büttner, Kunde, & Pfister, 2023; Galang, Cracco, Chirkov, Obhi, & Brass, 2024).

Given all of this, a number of limitations should be noted. First and foremost are the large number of participants that were dropped due to our exclusion criteria. This could potentially bias our sample towards participants who, ironically, are very good at following instructions (i.e., following rules). Furthermore, it should be noted that the effect sizes observed here are very small and required large sample sizes to become visible. While the complex nature of the experiment, especially regarding participant retention rates, made powering these experiments a challenge, future work should nevertheless attempt to better account for these factors. Relatedly, our data was completely collected online – as such, we had limited control over their environments during the experiment which may have added a significant amount of noise in the data (e.g., listening to music, watching a show, distracting background chatter, etc.). Another limitation of the current work is that rule breaking has no consequence or advantage – often one may choose to break a rule due to it leading to a shortcut (e.g., Pfister et al., 2019) which may be offset by a potential consequence if caught. However, in the current study, no such factors existed, and it was completely arbitrary for the participant to choose to follow or break the rule. As such, future work that uses a more nuanced design including these factors (e.g., a penalty for breaking the rule) will be important to shed more light on this topic. Finally, the fact that only the rule breaking effect at the 100 ms Delay is consistent across experiments warrants extreme caution in interpreting the effects found in the latter delays in Experiment 2.

CRediT authorship contribution statement

Carl Michael Galang: Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Conceptualization. **Ayça Akan:** Writing – review & editing, Software, Methodology, Investigation, Conceptualization. **Roland Pfister:** Writing – review & editing, Methodology, Conceptualization. **Marcel Brass:** Writing – review & editing, Supervision, Resources, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

CMG is supported by a Marie Skłodowska-Curie Actions (MSCA) Postdoctoral Fellowship (101105378). MB is supported by an Einstein Strategic Professorship from the Einstein Foundation (EPP-2018-483) and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2002/1 "Science of Intelligence" – project number 390523135. RP is funded in the Heisenberg Programme of DFG (PF853/10-1, 490925504).

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.concog.2025.103851.

Data availability

All data can be found on the following OSF page: https://osf.io/fnmc2/

References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Brown, V. A. (2021). An Introduction to Linear Mixed-Effects Modeling in R. *Advances in Methods and Practices in Psychological Science, 4*(1), 1–19. https://doi.org/10.1177/251524592096035

Chambon, V., & Haggard, P. (2012). Sense of control depends on fluency of action selection, not motor performance. *Cognition, 125*, 441–451. https://doi.org/10.1016/j.cognition.2012.07.011

Ciaunica, A., Atache, J., Haggard, P., Nakul, E., Bonnet, E., & Auvray, M. (2024). Explicit and implicit sense of agency in depersonalisation expériences. *Scientific Reports, 14*, 15396. https://doi.org/10.1038/s41598-024-65862-z

Demanet, J., Muhle-Karbe, P. S., Lynn, M. T., Blotenberg, I., & Brass, M. (2013). Power to the will: How exerting physical effort boosts the sense of agency. *Cognition, 129*, 574–578. https://doi.org/10.1016/j.cognition.2013.08.020

De Neys, W. (2006). Dual processing in reasoning: Two systems but one reasoner. *Psychological Science, 17*(5), 428–433. https://doi.org/10.1111/j.1467-9280.2006.01723.x

Ebert, J. P., & Wegner, D. M. (2010). Time warp: Authorship shapes the perceived timing of actions and events. *Consciousness and Cognition, 19*, 481–489. https://doi.org/10.1016/j.concog.2009.10.002

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods, 39*, 175–191. https://doi.org/10.3758/BF03193146

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*power 3.1: Tests for correlation and regression analyses. *Behav Res Methods, 41*, 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Galang, C. M., Malik, R., Kinley, I., & Obhi, S. S. (2021). Studying Sense of Agency Online: Testing the robustness of the intentional binding effect using an online sample. *Consciousness and Cognition, 95*, Article 103217. https://doi.org/10.1016/j.concog.2021.103217

Gozli, D. G. (2017). Behaviour versus performance: The veiled commitment of experimental psychology. *Theory & Psychology, 27*, 741–758. https://doi.org/10.1177/0959354317728130

Gozli, D. G. (2019). Experimental psychology and human agency. *Springer*. https://doi.org/10.1007/978-3-030-20422-8

Gutzeit, J., Weller, L., Kurten, J., & Huestegge, L. (2023). Intentional Binding: Merely a Procedural Confound? *Journal of Experimental Psychology: Human Perception and Performance, 49*(6), 759–773.

Haggard, P. (2017). Sense of agency in the human brain. *Nature Reviews Neuroscience, 18*, 197–208. https://doi.org/10.1038/nrn.2017.14

Haggard, P., Clark, S., & Kalogeras, J. (2002). Voluntary action and conscious awareness. *Nature Neuroscience, 5*(4), 382–385. https://doi.org/10.1038/nn827

Hoerl, C., Lorimer, S., McCormack, T., Lagnado, D. A., Blakey, E., Tecwen, E. C., & Buchner, M. J. (2020). Temporal Binding, Causation, and Agency: Developing a New Theoretical Framework. *Cognitive Science, 44*, Article e12843. https://doi.org/10.1111/cogs.12843

Kirsch, W., Kunde, W., & Herbort, O. (2019). Intentional Binding Is Unrelated to Action Intention. *Journal of Experimental Psychology: Human Perception and Performance, 45*(3), 378–385. https://doi.org/10.1037/xhp0000612

Klaffehn, A. A. L., Herbort, O., & Pfister, R. (2024). The fusion point of temporal binding: promises and perils of multisensory accounts. *Cognitive Psychology, 151*, Article 101662. https://doi.org/10.1016/j.cogpsych.2024.101662

Klaffehn, A., Sellmann, F. B., Kirsch, W., Kunde, W., & Pfister, R. (2021). Temporal binding as multisensory integration: Manipulating perceptual certainty of actions and their effects. *Attention, Perception, & Psychophysics, 83*, 3135–3145. https://doi.org/10.3758/s13414-021-02314-0

Kumle, L., Vo, M.-L.-H., & Draschkow, D. (2021). Estimating power in (generalized) linear mixed models: An open introduction and tutorial in R. *Behavior Research Methods, 53*, 2528–2543. https://doi.org/10.3758/s13428-021-01546-0

Langenberg, B., Janczyk, M., Koob, V., Kliegl, R., & Mayer, A. (2023). A tutorial on using the paired t-test for power calculations in repeated measures ANOVA with interactions. *Behavior Research Methods, 55*(5), 2467–2484. https://doi.org/10.3758/s13428-022-01902-8

Lange, K., Kuhn, S., & Filevich, E. (2015). "Just Another Tool for Online Studies" (JATOS): An Easy Solution for Setup and Management of Web Servers Supporting Online Studies. *PLoS ONE, 10*(6), Article e0130834. https://doi.org/10.1371/journal.pone.0134073

Lenth, R. V. (2024). emmeans: Estimated Marginal Means, aka Least-Squares Means. *R package version, 1*(9). https://doi.org/10.1080/00031305.1980.10483031

Ludecke, D. (2024). sjPlot: Data Visualization for Statistics in Social Science. *R package version, 2*(9), 7. https://doi.org/10.5281/zenodo.1308157

Magezi, D. A. (2015). Linear mixed-effects models for within-participant psychology experiments: An introductory tutorial and free, graphical user interface (LMMgui). *Frontiers in Psychology, 6*(2), 1–7. https://doi.org/10.3389/fpsyg.2015.00002

Malik, R., Galang, C. M., & Finger, E. (2022). The sense of agency for brain disorders: A comprehensive review and proposed framework. *Neuroscience and Biobehavioral Reviews, 139*, Article 574104759. https://doi.org/10.1016/j.neubiorev.2022.104759

Mariano, M., Stanco, G., Grapes, D. I., Rossetti, I., Bolognini, N., Paulesu, E., & Zapparoli, L. (2024). The sense of agency in near and far space. *Consciousness and Cognition, 120*, Article 103672. https://doi.org/10.1016/j.concog.2024.103672

Moore, J. W., & Obhi, S. S. (2012). Intentional binding and the sense of agency: A review. *Consciousness and Cognition, 21*, 546–561. https://doi.org/10.1016/j.concog.2011.12.002

Palan, S., & Schitter, C. (2018). Prolific.ac – A subject pool for online experiments. *Journal of Behavioral and Experimental Finance, 17*, 22–27. https://doi.org/10.1016/j.jbef.2017.12.004

Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology, 70*, 153–163. https://doi.org/10.1016/j.jesp.2017.01.006

Galang, C.M., Cracco, E., Chirkov, V., Obhi, S.S., & Brass, M. (2024). Sense of agency during Group Control. Preprint: https://osf.io/x76fk_v1.

Pfister, R. (2022). Operationalization and generalization in experimental psychology: A plea for bold claims. In: D. Gozli & J. Valsiner (Eds.), Experimental psychology: Ambitions and possibilities. DOI: 10.1007/978-3-031-17053-9_4.

Pfister, R., Wirth, R., Schwarz, K. A., Steinhauser, M., & Kunde, W. (2016). Burdens of non-conformity: Motor execution reveals cognitive conflict during deliberate rule violations. *Cognition, 147*, 93–99. https://doi.org/10.1016/j.cognition.2015.11.009

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. https://doi.org/10.1016/j.joep.2018.06.005

Schwarz, K. A., Tonn, S., Büttner, J., Kunde, W., & Pfister, R. (2023). Sense of agency in social hierarchies. *Journal of Experimental Psychology: General, 152*(5), 1013–1032. https://doi.org/10.1037/xge0001426

Schwarz, K. A., Weller, L., Klaffehn, A. L., & Pfister, R. (2019). The effects of action choice on temporal binding, agency ratings, and their correlation. *Consciousness and Cognition, 75*, Article 102807. https://doi.org/10.1016/j.concog.2019.102807

Schielzeth, H., Dingemanse, N. K., Nakagawa, S., Westnear, D. F., Allegue, H., Teplitsky, C., Reale, D., Dochtermann, N. A., Garamzergi, L. Z., & Araya-Ajoy, Y. G. (2020). Robustness of linear mixed-effects models to violations of distributional assumptions. *Methods in Ecology and Evolution, 11*, 1141–1152. https://doi.org/10.1111/2041-210X.13434

Suzuki, K., Lush, P., Seth, A. K., & Roseboom, W. (2019). Intentional Binding Without Intentional Action. *Psychological Science, 30*(6), 842–853. https://doi.org/10.1177/0956797619842191

Spaccasassi, C., Cenka, K., Petkovic, & Avenanti, A. (2023). Sense of agency predicts severity of moral judgments. *Frontiers in Psychology, 13*, Article 1070742. https://doi.org/10.3389/fpsyg.2022.1070742

Vastano, R., Pozzo, T., & Brass, M. (2017). The action congruency effect on the feelings of agency. *Consciousness and Cognition, 51*, 212–222. https://doi.org/10.1016/j.concog.2017.04.002

Weller, L., Schwarz, K. A., Kunde, W., & Pfister, R. (2020). Something from nothing: Agency for deliberate nonactions. *Cognition, 196*(104136), 1–10. https://doi.org/10.1016/j.cognition.2019.104136

Wiesing, M., & Zimmermann, E. (2024). Intentional binding – is it just causal binding? A replication study of Suzuki et al. (2019). *Consciousness and Cognition, 119*, Article 103556. https://doi.org/10.1016/j.concog.2024.103665

Wickham, H. (2016). ggplot2: Elegant graphics for data analysis. *Springer-Verlag New York.*. https://doi.org/10.1007/978-3-319-24277-4

Wirth, R., Foerster, A., Rendel, H., Kunde, W., & Pfister, R. (2018a). Rule-violations sensitise towards negative and authority-related stimuli. *Cognition and Emotion, 32*(3), 480–493. https://doi.org/10.1080/02699931.2017.1316706

Wirth, R., Foerster, A., Herbort, O., Kunde, W., & Pfister, R. (2018b). This is how to be a rule breaker. *Advances in Cognitive Psychology, 14*(1), 21–37. https://doi.org/10.5709/acp-0235-2

Wirth, R., Pfister, R., Foerster, A., Huestegge, L., & Kunde, W. (2016). Pushing the rules: Effects and aftereffects of deliberate rule violations. *Psychological Research, 80*, 838–852. https://doi.org/10.1007/s00426-015-0690-9

Wolpe, N., Haggard, P., Siebner, H. R., & Rowe, J. B. (2013). Cue integration and the perception of action in intentional binding. *Experimental Brain Research, 229*, 467–474. https://doi.org/10.1007/s00221-013-3419-2

Yeung, N., Botvinick, M. M., & Cohen, J. D. (2004). The neural basis of error detection: Conflict monitoring and the error-related negativity. *Psychological Review, 111*(4), 931–959. https://doi.org/10.1037/0033-295x.111.4.939

Zimmermann, E., & Cicchini, G. (2020). Temporal context affects interval timing at the perceptual level. *Scientific Reports, 10*(1), 8767. https://doi.org/10.1038/s41598-020-65609-6