Consciousness and Cognition 140 (2026) 104022

Contents lists available at ScienceDirect

Consciousness and Cognition

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

Full Length Article

Method matters: Diverging patterns in online measures of temporal binding

Crystal A. Silver [a,b], Benjamin W. Tatler [b], Ramakrishna Chakravarthi [b], Bert Timmermans [b,*]

[a] Department of Sociological and Psychological Sciences, Faculty of Social and Applied Sciences, Abertay University Dundee, Dundee DD1 1HG, UK
[b] School of Psychology, University of Aberdeen, Aberdeen AB24 3FX, UK

ARTICLE INFO

Keywords:
Temporal Binding
Online testing
Time perception
Sense of Agency

ABSTRACT

Temporal Binding (TB) refers to the subjective compression of time between an action and its outcome. While TB has been demonstrated across various paradigms, questions remain about its methodological reliability, particularly in online settings where technical control is limited. The present study examined whether two common interval judgement methods, estimation and replication, yield comparable TB effects when implemented online and whether either method is differentially sensitive to social contextual framing.

In Experiment 1, participants judged action-effect intervals using estimation or replication, across action-reaction and observation conditions. Results revealed that the two methods produced different and often opposing patterns of results across action-effect intervals.

Experiment 2 introduced a minimal social manipulation in which participants were led to believe that a previous participant generated interval outcomes. Again, method-specific patterns emerged, depending on different action-effect intervals.

These findings raise concerns about the robustness of TB effects as measured in online environments. The observed divergences across methods suggest that interval judgement paradigms may be highly sensitive to procedural and contextual factors. We suggest that TB effects obtained through online interval estimation or replication should be interpreted with caution, as the presence or absence of a TB effect may be due to the specific method used or the presented action-effect intervals.

1. Introduction

Temporal Binding refers to the perceived compression of time between a voluntary action and its sensory consequence (Borhani et al., 2017; Buehner, 2012a; Haggard et al., 2002; Majchrowicz & Wierzchoń, 2018). This compression is attributed to shifts in the perceived timing of the action (action shift), the outcome (outcome shift), or both (Libet, 2002; Tanaka et al., 2019). Unlike Intentional Binding, which presupposes that the action must be intentional, Temporal Binding requires only a perceived causal link between the action and its effect (Buehner, 2015; Moore & Obhi, 2012). This broader conceptualisation has supported its widespread use in studies of causality, perception and Sense of Agency across both social and nonsocial domains.

Despite its broad applicability, findings from Temporal Binding studies vary considerably. Some research reports stronger binding at shorter action-effect intervals (e.g., Haggard et al., 2002; Ruess et al., 2018), whereas other studies observed increased binding at longer intervals (e.g., Dewey & Knoblich, 2014; Humphreys & Buehner, 2009; Wen et al., 2015). These discrepancies raise concerns about the measurement reliability of Temporal Binding, potentially arising from methodological differences such as the specific intervals tested and the way perceived duration is measured (e.g., Ivanof et al., 2022; Siebertz & Jansen, 2022; Silver et al., 2025). For example, in Libet-style paradigms in which participants have to report where the dial of a clock was when they initiated an action or experienced an effect, clock speed has been shown to modulate binding (Ivanof et al., 2022), and action and outcome shifts have been found to be uncorrelated, suggesting different underlying mechanisms (Siebertz & Jansen, 2022; Tonn et al., 2021).

These concerns are especially relevant in online settings, where researchers have less control over factors such as screen refresh rates, input lag, and potential distractions in participants' environments (Anwyl-Irvine et al., 2021; Barbosa et al., 2023). Such variability presents particular challenges for studies investigating time-sensitive constructs like Temporal Binding, which depend on the precise perception and reporting of temporal intervals. Several studies have demonstrated that Temporal Binding effects can be reliably observed in uncontrolled

* Corresponding author at: School of Psychology, William Guild Building, University of Aberdeen, Aberdeen AB24 3FX, UK.
E-mail address: bert.timmermans@abdn.ac.uk (B. Timmermans).

https://doi.org/10.1016/j.concog.2026.104022
Received 22 September 2025; Received in revised form 16 January 2026; Accepted 11 February 2026
Available online 19 February 2026
1053-8100/

online environments, suggesting that web-based methods are viable (Cubillaset al., 2020; Flanagan et al., 2025; Galang et al., 2021; Le et al., 2026). However, they also highlighted that variability in hardware and software can influence the magnitude of binding effects, underscoring the importance of careful experimental design and timing control. Wider empirical evidence increasingly supports the viability of online methods for measuring a range of cognitive phenomena. Studies have shown that performance on tasks involving reaction time, attention, and temporal estimation is largely comparable between online and laboratory settings, provided that appropriate methodological safeguards are in place (Anwyl-Irvine et al., 2021; Germine et al., 2012). These safeguards include pre-screening for device compatibility, the use of browser-based calibration routines, implementing attention and compliance checks, and excluding participants based on timing irregularities or response latency distributions (Barbosa et al., 2023; Uittenhove et al., 2023). Although data collected online may show greater variability in timing precision due to uncontrolled hardware and environmental factors, effect sizes and internal consistency often remain comparable to those observed in laboratory settings (Germine et al., 2012; Uittenhove et al., 2023). Importantly, researchers may need to increase sample sizes to maintain statistical power when testing online, due to slightly elevated noise in response data (Barbosa et al., 2023). These findings suggest that while online platforms introduce specific challenges, well-designed studies can still yield reliable results. For paradigms such as Temporal Binding, which are highly sensitive to small perceptual differences in time, ongoing validation in web-based contexts remains essential.

The specific method used to assess perceived duration may, in itself, influence binding effects. Two widely used approaches are interval estimation and interval replication. Estimation involves reporting the perceived duration using a keyboard response, verbal report, or a visual analogue scale (Cravo et al., 2013; Imaizumi & Tanno, 2019; Vogel et al., 2021). Replication requires participants to reproduce the interval by initiating and terminating an action that matches the perceived duration (Stephenson et al., 2018). These methods impose different cognitive and motor demands. Replication may be affected by motor planning and execution constraints, especially for shorter intervals, which can artificially inflate binding effects (Gilden et al., 1995). Estimation, in contrast, can be performed without temporal pressure and may therefore be less susceptible to these confounds.

Despite these methodological differences, studies often compare Temporal Binding findings across estimation and replication paradigms without considering how the measurement method might influence results. This is particularly problematic in studies of Social Agency, where Temporal Binding effects are often interpreted in light of differences between social and nonsocial conditions. For instance, Vogel et al. (2021, 2022) reported enhanced binding in social conditions but noted that extremely short intervals may approach the perceived limits of human response speed, potentially constraining the magnitude of binding.

The impact of the above described factors on temporal binding may differentially manifest depending on the method used. Temporal Binding studies typically do not consider absolute amounts of subjective time compression, but compare the relative time compression between different conditions of interest. This means that between studies and sometimes within a study, it is not always easy or possible to assess the reliability of the method used and it is not always clear whether a Temporal Binding difference between conditions is a consequence of the method used, since usually only one method is being used in the first place.

The present study investigates whether estimation and replication measures produce comparable Temporal Binding effects when implemented online, and whether they do so consistently over a large number of action-effect intervals. In the first experiment, we addressed the typical comparison for Temporal Binding studies, comparing a condition in which a participant's action produces an effect, with a condition in which the participant observes an action and its effect. We tested whether estimation and replication yield similar Temporal Binding patterns across different action-effect intervals when intervals were merely observed or initiated. In the second experiment, building on previous research into Temporal Binding in social contexts (Pfister et al., 2014; Silver et al., 2025; Vogel et al., 2021; 2022), we assessed whether introducing a minimal social context, by presenting the effect as a prerecorded human response, produces a modulation of Temporal Binding and whether this effect depends on the measurement method.

The purpose of this research is not to determine which method is superior but rather to examine whether different measurement techniques yield consistent results or whether the methods interact with the target effects (observation/action or nonsocial/social). Observing differences in sensitivity between estimation and replication would suggest that online Temporal Binding effects are highly method-dependent. Such a finding would have implications for the interpretation of online Temporal Binding modulation.

2. General method

We conducted two experiments where the parameters of an interval judgement task were systematically manipulated. Experiment 1 examined replication and estimation measurement methods in an interval judgement task, comparing the relative sensitivity of each measurement to Temporal Binding effects expressed as the difference between executed action-effect intervals and observed action-effect intervals. Experiment 1 was entirely nonsocial. Experiment 2 investigated the sensitivity of each measurement to social-nonsocial differences using online protocols. It introduced a deceptive social element: participants in the social conditions believed that the response to their action or the observed interval was created by a previous participant. It did not include an observation (control) condition, as it was primarily interested in the social-nonsocial difference. Methods common to both experiments are described first, and the specifics of each experiment follow.

2.1. Materials and apparatus

The experiments were conducted online through the Testable behavioural research platform (Rezlescu et al., 2020; see Bridges et al., 2020 for timing accuracy compared to other platforms). Testable audience settings were restricted to those using desktop or laptop computers only, not phones or tablets. The operating system, web browser, and screen size were not controlled.

The stimuli used throughout the study were a schematic 'hand' above a 'button'. The 'hand' was a solid black hand icon ending at the wrist with only the index finger extended and pointing directly downwards, and all other fingers curled into the palm. The tip of the index finger always touched the top middle point of the 'button'. The 'button' was made from two coloured rectangles. The base of the 'button' was a short, wide black rectangle. The 'button' itself was a narrower, taller coloured rectangle, either green or blue contingent on trial sequence parameters (see Fig. 1). The blue 'button' was also depicted as 'pressed' by reducing the height of the rectangle by half.

2.2. Procedure

The experiments began with the Testable in-built calibration procedure. Then, participants were asked to read and complete on-screen instructions and demographic data questions (i.e., age, gender). They then completed 5 practice trials, followed by 60 experimental trials of whichever measurement method participants were assigned to. Experimental trials were presented with interval duration randomised. There were 2 break opportunities, after 20 and 40 trials. Practice trials followed the same stimuli/action sequence as experimental trials (see Fig. 1), and the intervals tested were sampled from the 7 in the main experiment; the shortest and longest intervals were always presented, and the other three were sampled at random from the rest of the

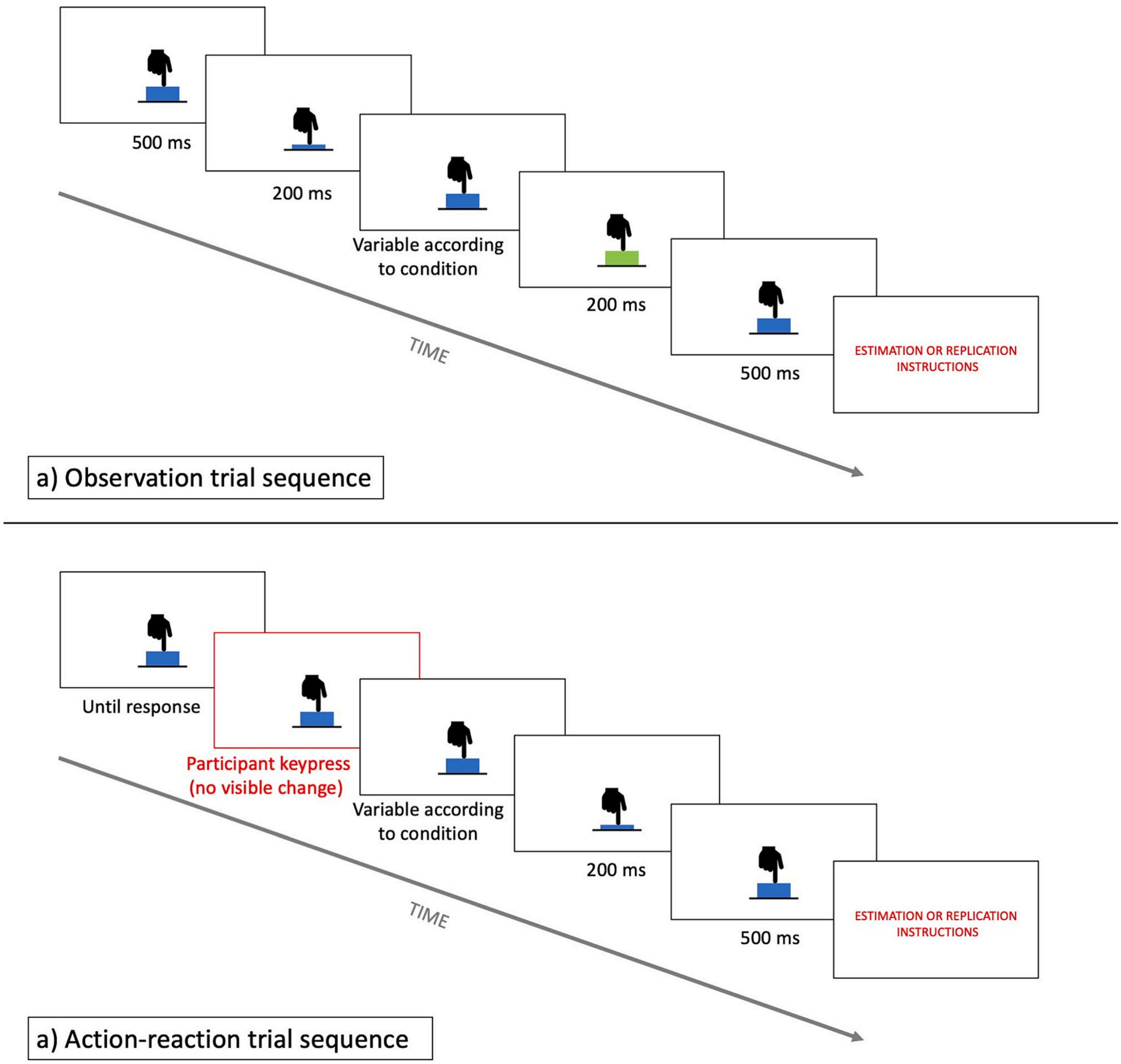

Fig. 1. Trial sequences for the observational (panel a) and action-reaction (panel b) conditions of the interval judgement paradigm.

durations. Demographic data were once again collected, and participants were asked an open-text box question to describe the task they had just completed; both served as attention checks (Buchanan & Scofield, 2018).

The observation trial sequence (Fig. 1, panel a) started with the hand and blue button presented at the centre of the screen. 500 ms later, the button was pressed by the hand for 200 ms. After a variable interval (200–2600 ms, 400 ms increments; 7 durations total), the button turned green for 200 ms, returning to blue for a further 500 ms. Finally, depending on the condition, the replication or estimation task instructions were displayed until response. Estimation instructions for the observation condition were 'Using the number keys, state (in milliseconds from 0−3000) how long you estimate the interval between the button push and green flash'. Replication instructions for the observation condition were 'Press the SPACE key twice, replicating the time interval between the button push and green flash'. The action-reaction trial sequence (panel b) started in the same way as the observation condition. Participants were asked to push the SPACE key (as per on-screen instructions prior to trial start) at a time of their choosing. After a variable interval (same intervals as the observation condition above) the blue on-screen button was 'pressed' for 200 ms, returning to its original position for a further 500 ms before the task instructions. For the action-reaction condition, replication instructions were 'Press the SPACE key twice, replicating the interval between your button push and the button push on-screen' for the action-reaction condition, and estimation instructions were 'Using the number keys, state (in milliseconds from 0−3000) how long you estimate the interval between your space key push and the button push on-screen'. Instructions were displayed until response.

3. Experiment 1

This experiment aimed to investigate whether replicating or estimating interval durations revealed different sensitivities to typical Temporal Binding effects, i.e., the difference between a condition where participants actively cause an effect, versus a condition where participants merely observe an action and an effect without active involvement (Buehner & Humphreys, 2009; Dewey & Knoblich, 2014; Haggard et al., 2002; Ruess et al., 2018).

3.1. Method

3.1.1. Participants

210 participants (134 females, age 18–32 years; mean age = 23.2) were recruited either through Prolific (Prolific.com) or from among psychology undergraduate students registered on the University of Aberdeen's SONA Systems. Participants by condition: estimation/action-reaction = 47, replication/action-reaction = 58, estimation/observation = 56, replication/observation = 49. Sample size was determined as roughly 40 participants for each of the 4 groups (see 3.1.2 Design). The Method × Action effect we wanted to investigate was not studied before, nor did we, in fact anticipate an effect of any particular size. Taking into account flaws of on-line data collection, we aimed for 4x50 = 200 participants.

Participants' data were rejected if they responded incorrectly on more than one attention check question (7 participants rejected prior to reaching the 210 participants above). Participants were either reimbursed or received course credit for their time. Written informed consent was obtained. The study was approved by the Psychology Ethics Committee, University of Aberdeen.

3.1.2. Design

The experiment used a 2 × 2 × 7 mixed design with two between-subjects factors of 'Method' (replication/estimation) and 'Action' (action-reaction/observation), and one within-subjects factor of 'Interval Duration' (200–2600 ms in 400 ms increments; 7 conditions total). Method and Action were tested in independent samples to keep the experiment within a reasonable duration for an online experiment and to avoid contamination between methods.

3.2. Statistical analysis

3.2.1. Outlier detection and exclusion criteria for participant-wise and trial-wise measures

First, any participant whose median response was less than 500 ms was excluded (39 participants excluded; 171 participants remaining). This hard criterion was implemented first to prevent participants not actively performing the task from skewing the other exclusion criteria. Then, participants for whom the correlation coefficient between the to-be-judged intervals and the corresponding interval estimations or replications were more than 2.5 absolute deviations from the group median correlation coefficient were excluded (median correlation coefficient for estimation = 0.856, for replication = 0.851; MAD; $b = 1.4826$; Leys et al., 2013) (31 participants excluded; 140 participants remaining). This ensured that participants included in the analysis had engaged with the task. Next, individual trials where the percentage response (i.e., estimate/interval × 100) was more than 2.5 absolute deviations from the median percentage response across all trials (median percentage response = 101%) were excluded (MAD; $b = 1.4826$; Leys et al., 2013) (8.1% of trials excluded). This conservatively removed response outliers. Finally, if any participant had more than 25% of their trials removed by the trial-wise criterion, they were excluded entirely from analyses (a further 4 participants were excluded). Participants included in the analysis after all exclusions were: 136 (96 females; 18 – 30 years; mean age = 24.4). Remaining participants after exclusions by condition: estimation/action-reaction = 26, replication/action-reaction = 35, estimation/observation = 38, replication/observation = 37.

3.2.2. Mixed linear Modelling: Fixed and random effect structure

Data were analysed using Mixed Linear Modelling (MLM; Kliegl et al., 2012). The fixed effects included in the MLM were as follows: categorical fixed effects of '*Method*' and '*Action*' with Helmert contrast coding, and a continuous fixed effect of *'Interval Duration'* (200–2600 ms), which was centred and scaled using the base *R scale* function ('*Interval_Scaled';* R Core Team, 2019) and wrapped in the *poly* function from the *stats* package (R Core Team, 2019), which fits polynomials up to degree 2 to scrutinise both linear and quadratic effects. *'Participant'* was included in the MLM as a random effect; as 'interval' was a within-subject factor, it was important to include Participant to account for any individual differences within groups. The MLM was built using the *lmer* function of the *lme4* package (Bates, Mächler, Bolker, & Walker, 2015) in the *R Studio* integrated development environment (Posit team, 2025); *ggplot2* (Wickham, 2016) was used for plotting data.

The maximal MLM was used for analysis (Barr et al., 2013). *Lmer* (Bates et al., 2015) syntax of the maximal model is shown below:

*lmer (Percentage_Binding ~ 1 + poly(Interval_Scaled, 2) * Method * Action + (1 + Interval_Scaled | Participant), control = lmerControl (optimizer = "bobyqa"))*

p-values for fixed effects within the MLM were generated using the *summary* function of base *R* (R Core Team, 2019) using Satterthwaite's method.

3.3. Results

This experiment aimed to investigate whether the two methods used to assess Temporal Binding differed in their ability to detect differences between a condition where participants actively cause an effect, and a condition where participants merely observe an action and an effect without active involvement. As such, the interaction between the *Method* and *Action* variables is of crucial interest, potentially depending on *Interval*.

The results are depicted in Fig. 2, and the output of the MLM is shown in Table 1. There was a linear main effect of Interval, which was qualified by an interaction with Method, with replication showing a downward trend, and estimation an upward trend as interval durations increased. There was also a quadratic Interval × Method interaction, with both replication and estimation slopes flattening out towards longer intervals, but mirroring each other. There was a quadratic Interval × Action interaction, with action-reaction curves starting steeper and curving more than observation curves.

Crucially, there was a quadratic Interval × Method × Action interaction, suggesting that across various intervals, the observed difference between action-reaction and observation groups depends on the method through which they assessed interval duration.

Follow-up MLMs were run on the replication and estimation conditions separately to establish if they exhibited linear and/or quadratic trends across intervals. The replication condition showed both a linear, $\beta = -1175.300$, $SE = 60.51$, $t = -19.42$, $p = 0 < .001$, and positive quadratic, $\beta = 406.31$, $SE = 26.69$, $t = 15.23$, $p = 0 < .001$, main effect of Interval on replications. The estimation condition showed both a linear $\beta = 228.943$, $SE = 109.17$, $t = 61.23$, $p = 0.040$, and negative quadratic, $\beta = -437.88$, $SE = 34.51$, $t = -12.69$, $p = 0 < .001$, main effect of Interval on estimations.

Of interest for our core hypothesis, we carried out follow-up 2x2 Analyses of Variance (ANOVAs) for the variables *Method* and *Action* for each interval duration to investigate the 3-way Interval × Method × Action interaction; significance was adjusted for 7 comparisons using the Bonferroni-Holm method (Holm, 1979). Table 2 lists the Method × Action interactions, which were significant at the 200, 1000, 1400, 1800 and 2200 ms interval durations. In addition, there was a main effect of Action in the 1400 (Sum Sq = 11502, F Value = 9.60, p = 0.002), 1800 (Sum Sq = 4789, F Value = 5.39, p = 0.020) and 2600 ms (Sum Sq = 2552, F Value = 5.22, p = 0.023) interval durations.

3.4. Discussion

Experiment 1 showed that, at least in an online setting, detecting Temporal Binding effects (i.e., action-reaction vs observation condition differences, henceforth denoted 'Action') depended on the measurement method used (i.e., interval replication vs estimation) and varied across

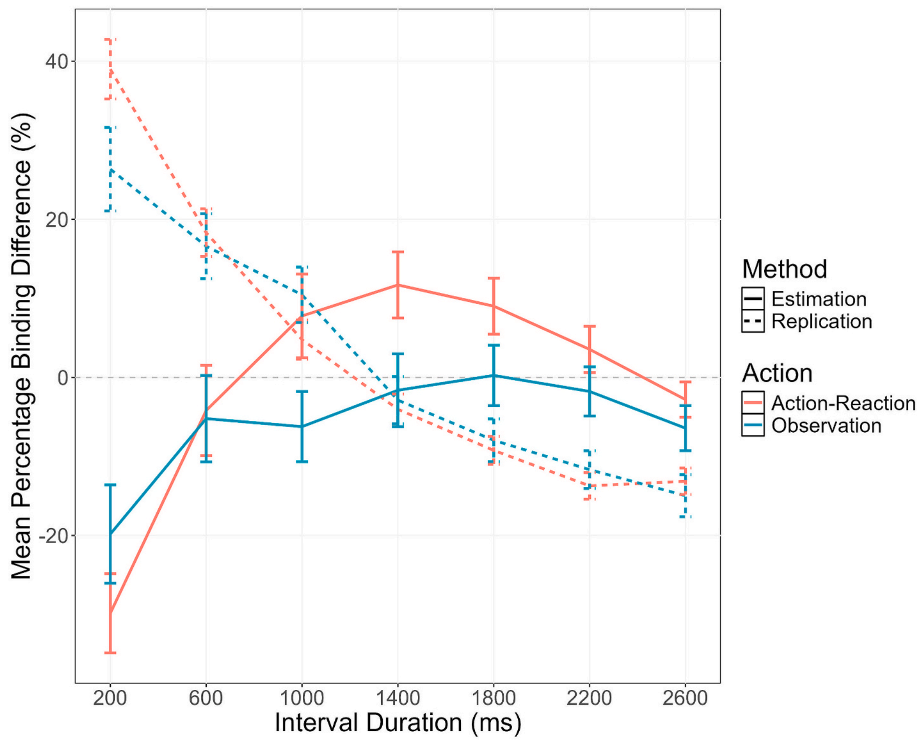

Fig. 2. Temporal Binding (%change relative to presented intervals) plotted as a function of interval for estimation (solid line), replication (dashed line), action-reaction (red) and observation (green) conditions. Error bars represent 95% confidence intervals. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Results of the Mixed Linear Model to predict Temporal Binding in Experiment 1. Significant predictors are highlighted in bold.

Fixed Effect	*b*	*SE*	*df*	*t*	*p*
Intercept	0.68	1.65	131.52	0.41	0.680
Interval (linear)	**−665.86**	**86.49**	**131.80**	**−7.70**	**<0.001**
Interval (quadratic)	−33.81	30.76	10813.71	−1.10	0.272
Method	−3.52	3.30	131.52	−1.07	0.288
Action	3.00	3.30	131.52	0.91	0.365
Interval (linear) × Method	**1962.28**	**172.98**	**131.80**	**11.34**	**<0.001**
Interval (quadratic) × Method	**−1202.07**	**61.52**	**10813.71**	**−19.54**	**<0.001**
Interval (linear) × Action	38.00	172.98	131.80	0.22	0.826
Interval (quadratic) × Action	**−132.51**	**61.52**	**10813.71**	**−2.15**	**0.031**
Method × Action	5.66	6.60	131.52	0.86	0.393
Interval (linear) × Method × Action	476.60	345.96	131.80	1.38	0.171
Interval (quadratic) × Method × Action	**−1065.28**	**123.04**	**10813.71**	**−8.66**	**<0.001**
Random Effect		Element	Variance	SD	Corr.
Participant		Intercept	350.4	18.72	
		Interval	78.9	8.88	−0.40

Note: to check for robustness, the above analysis was also executed with different exclusion criteria (median = 400 ms and 3 × MAD), which yielded comparable outcomes.

Table 2
Results of the follow-up ANOVAs scrutinising the Method × Action interaction for each interval duration. Bonferroni-Holm adjusted significance is denoted by bold font.

Interval	Sum Sq	F Value	p
200	**34346**	**17.74**	**<0.001**
600	44	0.02	0.887
1000	**38,390**	**23.07**	**<0.001**
1400	**20552**	**17.15**	**<0.001**
1800	**9941**	**11.20**	**<0.001**
2200	**5354**	**8.27**	**0.004**
2600	267	0.55	0.460

action-effect interval durations. As it cannot be concluded which measurement method's sensitivity is more reliable, this brings into question the online use of the Temporal Binding method in general. Analyses of the effect of Action within each method showed that while estimation was sensitive to the Action manipulation at the midmost intervals (1000, 1400, 1800 ms), for replication, in contrast, this was only observed at 200 ms.

Further caution is needed as all these effects were in the opposite direction from what was expected: observation eliciting *shorter* interval judgements than action-reaction. One would expect intervals to be judged as longer in the observation condition, compared with the action-reaction condition. The most obvious reason for this is that the to-be-judged interval in the observation condition was inherently shorter than that of the action-reaction condition. When observing an instant on-screen action initiation (i.e., single frame stimulus change), followed by an instant on-screen reaction, there is no ambiguity as to when the start and end of the interval is. In contrast, in the action condition, the participant initiates their hand and finger movement to perform the key press before the actual interval trigger is recorded (i.e., the button press). As a consequence, exactly when participants perceive the interval as having started is necessarily fuzzy and unknown. It is plausible that participants perceive the interval as having commenced when they decided to act, when they initiated movement, or when they pressed the button. The first two would make the perceived interval longer than the recorded interval and work against any recorded action-reaction time compression. In hindsight, a more appropriate observation condition would have had a button press animation start the action-reaction interval, introducing movement and therefore an extended period of time before the moment the on-screen button is pressed.

Despite this flaw, and hence the absence of typical Temporal Binding ("Action") effects in the predicted direction, what is important is that the same observation condition was used for all methods and intervals, and hence the conclusion holds that the observed Action effect depends crucially on the method used to evaluate time intervals. Even if one assumes that both methods' observation groups would, given an extended interval initiation such as a hand movement animation, judge those intervals as longer, the result might be an absence of Action effect with estimation and a clear effect of Action (in the expected direction) with replication, which would still reflect an important disparity between methods. Still, it cannot be ruled out that a different, more appropriate control condition might have a different effect depending on the method used. For example, showing a longer video to initiate the action could lessen differences in interval estimation while leaving replication largely unaffected, or affecting it to a smaller extent. As the results diverge widely between methods, it is impossible to tell.

There was a pronounced over-replication of the shortest interval (200 ms), which was also reported by Silver et al. (2025), where replication responses to short intervals were similarly inflated. Such over-replication is likely due to the difficulty in accurately reproducing such brief durations. Replication involves not just perceptual evaluation but also the initiation and termination of a motor response, which introduces a fixed motor delay (Gilden et al., 1995). This delay disproportionately affects shorter intervals and may reflect interference with internal timing mechanisms such as the Internal Clock (a neural pacing signal), which is sensitive to both stimulation and motor activity (Wenke and Haggard, 2009). Such over-replication could alternatively, or in addition, be due to Vierordt's Law, leading to over-replicating shortest intervals and under-replicating longest intervals in any given duration range (Glasauer & Shi, 2021) Overall, as durations increased, replications shortened in comparison to the to-be-judged interval. As predicted, this trend was not found when intervals were estimated. Estimations were shorter than the to-be-judged intervals at either end of the interval range (200, 600 and 2600 ms) and longer than the to-be-judged intervals 1400 – 1800 ms. Temporal Binding studies in extant literature rarely test such a comprehensive range of action-effect intervals (there are exceptions, see Humphreys & Buehner, 2009). Tentatively, the trend for interval replication is in line with Dewey and Knoblich (2014). Participants in that study replicated intervals of 200, 400 and 1200 ms; replication durations proportionally were compressed more as interval duration increased. Conversely, the trend for interval estimation contradicts those of extant literature. Humphreys and Buehner (2009) also report nonsocial Temporal Binding effects that increase as interval durations increase with comparable durations to the current study (0 – 2000 ms) using an estimation method, and Vogel et al. (2021, 2022) report proportionally shorter estimations at 700 ms than 400 ms, for social and nonsocial conditions alike. In contrast, estimation results show that between 200 ms and 1400 ms, there is progressively less time compression, whereas over all durations, there is no linear trend.

4. Experiment 2

Experiment 2 was similar to Experiment 1 but investigated whether different measurement methods exhibited different sensitivity to a conservative social-nonsocial context manipulation, which typically yields stronger Temporal Binding in the social context (Silver et al., 2025; Vogel et al., 2021; 2022). To this end, a 'social' condition was introduced where participants were told that the reactions to their actions were pre-recorded from a previous participant. This new conservatively social condition would be analysed against the nonsocial action-reaction condition from Experiment 1. As results from Experiment 1 demonstrated that the observation condition was potentially flawed and hence the effect of *Action* difficult to interpret, the decision was made to exclude the observation condition from Experiment 2. This allowed us, as in Silver et al. (2025) Experiment 1, to focus solely on the relative comparison of social vs nonsocial conditions across intervals, as in Experiment 1.

4.1. Method

4.1.1. Participants

206 participants (112 females, age 18–38 years; mean age = 22.1) were recruited from among psychology undergraduate students registered on the University of Aberdeen's SONA Systems. Participants by condition: nonsocial/estimation = 47, nonsocial/replication = 58, social/estimation = 57, social/replication = 44. This experiment was run in parallel to Experiment 1, hence we could not use Experiment 1′s effect size to guide sample size. Instead, we applied the same logic as for Experiment 1, aiming for roughly 4 × 50 = 200 participants.

Participants' data were excluded if they responded incorrectly on more than one attention check question (4 participants were rejected prior to reaching the 206 participants above). Participants were reimbursed monetarily or with course credits for their time. Written informed consent was obtained from all participants. The study was approved by the Psychology Ethics Committee, University of Aberdeen.

4.1.2. Design

Experiment 2 used a 2 × 2 × 7 mixed design with two between-subjects factors of 'Socialness' (social/nonsocial) and 'Method' (estimation/replication), and one within-subjects factor of 'Interval Duration' (200–2600 ms in 400 ms increments; 7 conditions total).

4.1.3. Procedure

The procedure was the same as described in General Methods, with one additional element: in the social condition, participants were told that the reactions to their actions were pre-recorded from a previous participant. They were informed that the previous participant had been instructed to react to a 'go' signal at a time of their choosing. This information was included in the task instructions before practice trials.

4.2. Statistical analysis

4.2.1. Outlier detection and exclusion criteria

Outlier detection and exclusion criteria were as in Experiment 1. 42 participants were excluded for having median responses less than 500 ms. 30 further participants were excluded because their interval/response correlation coefficients were more than 2.5 times the MAD condition correlation coefficient (median correlation coefficient, estimation = 0.863, replication = 0.849). 7.4% of trials were removed because their percentage response (estimate/interval x 100) was more than 2.5 times the MAD of the overall median percentage response (median percentage response = 103%). Finally, another 5 participants were excluded as more than 25% of their trials had to be excluded using the trial-wise exclusion criterion mentioned above. One hundred and twenty-nine participants (83 females; 18 – 36 years; mean age = 22.1) were included in the analysis after all exclusions were applied in Experiment 2. Remaining participants after exclusions by condition: nonsocial/estimation = 26, nonsocial/replication = 35, social/estimation = 34, social/replication = 34.

4.2.2. MLM: Fixed and random effect structure

The fixed effects included in the MLM were as follows: categorical fixed effects of '*Socialness*' and '*Method*' with Helmert contrast coding, and a continuous fixed effect of '*Interval Duration*' (200–2600 ms), which was centred and scaled and wrapped in the *poly* function, which fits polynomials up to degree 2 to scrutinise both linear and quadratic effects. '*Participant*' was included in the MLM as a random effect, as in Experiment 1.

The maximal MLM was used for analysis (Barr et al., 2013). *lmer* (Bates et al., 2015) syntax of the maximal model for Experiment 2 is shown below:

*lmer (Percentage_Binding ~ 1 + poly(Interval_Scaled, 2) * Socialness * Method + (1 + Interval_Scaled | Participant), control = lmerControl (optimizer = "bobyqa"))*

4.3. Results

This experiment aimed to investigate whether the methods used to assess Temporal Binding are differentially sensitive to the socialness effect. As such, a Method × Socialness interaction is of crucial interest, potentially depending on Interval.

The results are depicted in Fig. 3 (data are available at https://osf.io/ahf4q/), and the output of the MLM is shown in Table 3. There was a linear main effect of Interval on Temporal Binding, which was qualified by an interaction with Method, replication showing a downward trend, estimation not. There was also a quadratic Interval × Method interaction, with the replication slope curving and flattening out towards longer intervals, curving from lesser to greater binding, estimation curving quadratically, with the least binding at midmost intervals. Crucially, there was a quadratic Interval × Method × Socialness 3-way interaction. This suggests that across various intervals, the observed difference between social and nonsocial groups depends on the interval judgement method used.

Follow-up MLMs were run on the replication and estimation conditions separately to establish if they exhibited linear and/or quadratic trends across intervals. The replication condition had both a linear, β = *-1195.98, SE = 61.15, t = -19.56, p = 0<.001*, and quadratic, β = *424.90, SE = 27.34, t = 15.54, p = 0<.001*, main effect of Interval. The estimation condition had an inversely signed quadratic, β = *-465.47, SE = 36.44, t = -12.77, p = 0<.001*, main effect of Interval.

Table 3
Results of the Mixed Linear Model to predict Temporal Binding in Experiment 2. Significant predictors are highlighted in bold.

Fixed Effect	b	SE	df	t	p
Intercept	1.68	1.83	124.55	0.91	0.363
Interval (linear)	**−729.85**	**92.05**	**124.82**	**−7.93**	**<0.001**
Interval (quadratic)	−42.81	32.08	10447.89	−1.34	0.182
Method	−1.72	3.67	124.55	−0.47	0.640
Socialness	1.07	3.67	1242.55	0.29	0.772
Interval (linear) × Method	**1903.10**	**184.10**	**124.82**	**10.34**	**<0.001**
Interval (quadratic) × Method	**−1270.61**	**64.17**	**10447.89**	**−19.80**	**<0.001**
Interval (linear) × Socialness	175.70	184.10	124.82	0.95	0.342
Interval (quadratic) × Socialness	−122.01	64.16	10447.89	−1.75	0.081
Method × Socialness	1.27	7.33	124.55	0.17	0.863
Interval (linear) × Method	518.60	368.20	124.82	1.41	0.162
Interval (quadratic) × Method × Socialness	**−904.42**	**128.33**	**10447.89**	**−7.05**	**<0.001**
Random Effect		Element	Variance	SD	Corr.
Participant		Intercept	414.92	20.37	
		Interval	88.83	9.43	−0.54

Note: to check for robustness, the above analysis was also executed with different exclusion criteria (median = 400 ms and 3 × MAD), which yielded comparable outcomes.

Follow-up 2x2 Analyses of Variance (ANOVAs) for the variables *Method* and *Socialness* were carried out for each interval duration to investigate the 3-way Interval × Method × Socialness interaction;

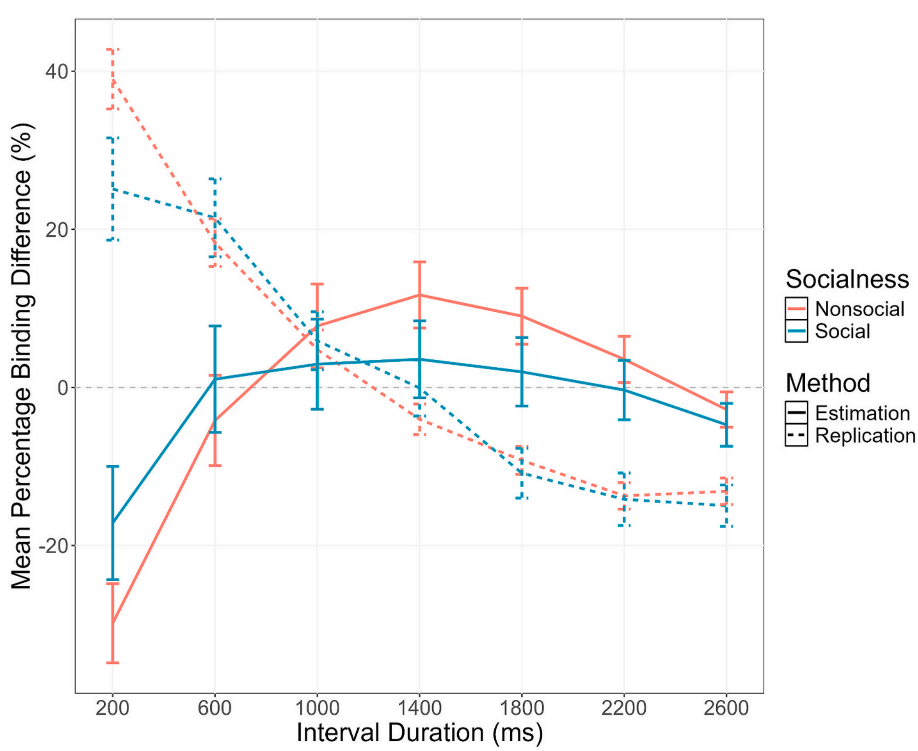

Fig. 3. Temporal Binding (% binding difference) plotted as a function of interval for estimation (solid line), replication (dashed line), nonsocial (red) and social (green) conditions. Error bars represent 95% confidence intervals. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

significance was adjusted for 7 comparisons using the Bonferroni-Holm method (Holm, 1979). Table 4 lists the Method × Socialness interactions, which were significant at the 200 and 1400 ms interval durations. Of note, in the follow-up ANOVA for 1800 ms there was a main effect of Socialness, Sum Sq = 6089, F Value = 6.78, p = 0.009. This was the only main effect of Socialness across interval duration ANOVAs.

4.4. Discussion

Experiment 2 showed that reliably detecting social-nonsocial differences in Temporal Binding effects differed between measurement methods and varied across action-effect interval durations. Measurement methods were differentially sensitive to the social-nonsocial differences at 200 and 1400 ms interval durations. While at durations of 1000 ms and longer, the estimated intervals were numerically lower in the social than in the nonsocial condition (in line with Vogel et al., 2022; Vogel et al., 2021), conservative follow-up tests revealed that there was only a main effect of Socialness at 1800 ms.

A key limitation in interpreting these findings lies in the nature of the social manipulation used. Although the experiments were intentionally conducted online to evaluate the feasibility of detecting Temporal Binding effects in uncontrolled settings, the implementation of the social condition relied solely on a cover story. Participants were informed that the outcomes were pre-recorded responses from previous participants, but no direct or interactive social engagement occurred. As such, the manipulation should be considered a minimal or conservative test of social context. That method-specific differences were still observed suggests some sensitivity to this framing, but these results should not be taken as definitive evidence for or against a Socialness effect. Rather, the findings highlight the difficulty of isolating genuine binding effects from methodological influences in online paradigms, particularly when measuring perceived interval durations using different techniques.

5. General discussion

Two experiments investigated whether Temporal Binding effects can be robustly detected in online environments and whether these effects depend on the measurement method used. Across both experiments, estimation and replication methods revealed diverging patterns of Temporal Binding, with effects varying in both magnitude and direction depending on the interval duration and the manipulation. This divergence was not merely a matter of degree but included contradictory patterns.

In the first experiment, differences in measurement method led to divergent observed differences between conditions in which participants initiated an action, versus where they only observed an action, depending on interval duration (e.g., Hon, 2023; Siebertz & Jansen, 2022). The degree to which estimation or replication detected action-observation differences, diverged at all action-effect intervals apart from 600 and 2600 ms. In general, replication responses became increasingly shorter relative to the to-be-judged intervals as durations increased, while estimation responses showed a non-linear pattern, with overestimations at midrange durations and underestimations at the shortest and longest intervals.

Table 4
Results of the follow-up ANOVAs scrutinising the Method × Socialness interaction for each interval duration. Bonferroni-Holm adjusted significance denoted by bold font.

Interval	Sum Sq	F Value	p
200	**46083**	**21.15**	**<0.001**
600	387	0.16	0.688
1000	3339	1.82	0.177
1400	**13966**	**10.92**	**<0.001**
1800	2551	2.84	0.092
2200	1032	1.43	0.232
2600	1	0.00	0.967

The second experiment tested whether a minimal social context (i.e., instruction that action-effect intervals were recorded from another person) modulated Temporal Binding (in line with Galang et. al., 2021). Results again differed between methods and intervals. The degree to which estimation or replication detected social-nonsocial differences, diverged at 200 and 1400 ms Estimation did show numerically lower judgements in the social condition at intervals of 1000 ms and longer, in line with previous work (Vogel et al., 2022; Vogel et al., 2021).

These findings raise methodological considerations. While both estimation and replication detected differences between action and observation conditions in Experiment 1, the timing and direction of these effects varied across virtually all intervals. For Estimation, observation intervals were generally judged as shorter, even though standard Temporal Binding findings would predict the opposite pattern (Buehner, 2012b). A likely explanation lies in the perceptual clarity of the interval onset in observation trials, which began and ended with discrete visual cues. In contrast, action trials may have introduced ambiguity around when the interval began, whether it was the intention to act, the motor initiation, or the actual button press (Libet, 2002; Tanaka et al., 2019). Such ambiguity could elongate perceived durations in action conditions, counteracting any expected binding. Introducing a visual cue at the onset of observation intervals might have helped match the temporal structure more closely across conditions. That said, the core issue remains the diverging sensitivity of the two measurement methods. Estimation may be more susceptible to perceptual or cognitive anchoring, while replication may rely on different attentional or memory-based processes. These differences complicate direct comparisons and highlight the need for caution when interpreting binding effects across methodological approaches.

These findings were complicated by the nature of the online setting. The social condition was minimal and entirely cover-story based, and no genuine social interaction or observation occurred. The fact that any pattern emerged at all suggests a potential sensitivity to this framing but does not clearly establish the reliability of a social effect. The experiment thus offers only tentative support for a social modulation of Temporal Binding. The minimal nature of the social context, coupled with the divergent outcomes across estimation and replication, makes it difficult to determine whether the observed effects reflect genuine social modulation or methodological artefacts.

Previous work has suggested that Temporal Binding effects can be detected in online environments (Cubillaset al., 2020; Flanagan et al., 2025; Le et al., 2026), but often using different paradigms such as the Libet clock method, which separates action and effect judgements and provides perceptual anchors (Galang et al., 2021). The present findings show that when estimation or replication methods are used in a standard interval judgement paradigm, Temporal Binding effects are not only inconsistent across intervals and manipulations but may also manifest in reversed or ambiguous directions depending on the method.

Overall, these experiments highlight fundamental challenges in reliably measuring Temporal Binding effects in online environments. The substantial divergence between estimation and replication methods (both in detecting effects and in the direction of those effects) raises serious concerns about the validity and reliability of online Temporal Binding paradigms. This variability suggests that Temporal Binding, as assessed through these interval judgement tasks, may be highly sensitive to methodological factors unrelated to the underlying perceptual phenomenon. Consequently, findings from online Temporal Binding studies using different measurement methods should be interpreted with caution, as the results may reflect measurement artefacts rather than genuine temporal distortions. Until there is clear evidence demonstrating consistent, replicable Temporal Binding effects across methods in online settings, claims about the presence or modulation of Temporal Binding in such contexts remain provisional at best. This underscores the urgent need for rigorous validation of online Temporal Binding paradigms before they can be confidently employed to study temporal perception or agency-related phenomena.

CRediT authorship contribution statement

Crystal A. Silver: Writing – review & editing, Writing – original draft, Visualization, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Benjamin W. Tatler:** Writing – review & editing, Supervision, Conceptualization. **Ramakrishna Chakravarthi:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization. **Bert Timmermans:** Writing – review & editing, Supervision, Resources, Methodology, Funding acquisition, Formal analysis, Conceptualization.

Funding

This work was supported by UKRI Biotechnology and Biological Sciences Research Council (BBSRC) grant BB/M010996/1 ("Mechanisms of Social Agency") to Crystal Silver, and a Carnegie Trust Research Incentive Grant RIG008270 ("Experiencing myself through you: Self-agency in social interaction") to Bert Timmermans and Ramakrishna Chakravarthi.

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.

References

Anwyl-Irvine, A. L., Dalmaijer, E. S., Hodges, N., & Evershed, J. K. (2021). Realistic precision and accuracy of online experiment platforms, web browsers, and devices. *Behavior Research Methods, 53*(4), 1407–1425. https://doi.org/10.3758/s13428-020-01501-5

Barbosa, J., Stein, H., Zorowitz, S., Niv, Y., Summerfield, C., Soto-Faraco, S., & Hyafil, A. (2023). A practical guide for studying human behavior in the lab. *Behavior Research Methods, 55*(1), 58–76. https://doi.org/10.3758/s13428-022-01793-9

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1). https://doi.org/10.18637/jss.v067.i01.

Borhani, K., Beck, B., & Haggard, P. (2017). Choosing, doing, and controlling: Implicit sense of agency over somatosensory events. *Psychological Science, 28*(7), 882–893. https://doi.org/10.1177/0956797617697693

Bridges, D., Pitiot, A., MacAskill, M. R., & Peirce, J. W. (2020). The timing mega-study: Comparing a range of experiment generators, both lab-based and online. *PeerJ, 8*, e9414.

Buchanan, E. M., & Scofield, J. E. (2018). Methods to detect low quality data and its implication for psychological research. *Behavior Research Methods, 50*(6), 2586–2596. https://doi.org/10.3758/s13428-018-1035-6

Buehner, M. J. (2012a). Temporal Binding. In *Attention and Time* (p. 14). Oxford University Press.

Buehner, M. J. (2012b). Understanding the past, predicting the future: Causation, not intentional action, is the root of temporal binding. *Psychological Science, 23*(12), 1490–1497. https://doi.org/10.1177/0956797612444612

Buehner, M. J. (2015). Awareness of voluntary and involuntary causal actions and their outcomes. *Psychology of Consciousness: Theory, Research, and Practice, 2*(3), 237–252. https://doi.org/10.1037/cns0000068

Buehner, M. J., & Humphreys, G. R. (2009). Causal Binding of Actions to their Effects. *Psychological Science, 20*(10), 1211–1228.

Cravo, A. M., Haddad, H., Claessens, P. M. E., & Baldo, M. V. C. (2013). Bias and learning in temporal binding: Intervals between actions and outcomes are compressed by prior bias. *Consciousness and Cognition*. https://doi.org/10.1016/j.concog.2013.08.001

Cubillas, C. P., Landáburu, Í., & Matute, H. (2020). Methodological factors involved in the study of temporal binding using the open source software labclock web. *Frontiers in Psychology, 11*, 1040. https://doi.org/10.3389/fpsyg.2020.01040

Dewey, J. A., & Knoblich, G. (2014). Do implicit and explicit measures of the sense of agency measure the same thing? *PLoS One1, 9*(10). https://doi.org/10.1371/journal.pone.0110118

Flanagan, L., Zumbrunn, N. M., Hirst, R. J., & McGovern, D. P. (2025). Assessing the reliability of an online measure of the temporal binding window of audiovisual integration. *Behavior Research Methods 57*, 261 (2025). https://doi.org/10.3758/s13428-025-02791-3.

Galang, C. M., Malik, R., Kinley, I., & Obhi, S. S. (2021). Studying sense of agency online: Can intentional binding be observed in uncontrolled online settings? *Consciousness and Cognition, 95*. https://doi.org/10.1016/j.concog.2021.103217

Germine, L., Nakayama, K., Duchaine, B. C., Chabris, C. F., Chatterjee, G., & Wilmer, J. B. (2012). Is the Web as good as the lab? Comparable performance from Web and lab in cognitive/perceptual experiments. *Psychonomic bulletin & review, 19*(5), 847–857. https://doi.org/10.3758/s13423-012-0296-9

Gilden, D. L., Thornton, T., & Mallon, M. W. (1995). 1/f Noise in Human Cognition. *Science*, 1837–1839.

Glasauer, S., & Shi, Z. (2021). The origin of Vierordt's law: The experimental protocol matters. *PsyCh Journal, 10*(5), 732–741. https://doi.org/10.1002/pchj.464

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

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, 6*(2).

Hon, N. (2023). Attention and expectation likely underlie temporal binding measured using the Libet Clock. *Quarterly Journal of Experimental Psychology, 76*(9), 2084–2093. https://doi.org/10.1177/17470218221132762

Humphreys, G. R., & Buehner, M. J. (2009). Magnitude estimation reveals temporal binding at super-second intervals. *Journal of Experimental Psychology: Human Perception and Performance, 35*(5), 1542–1549. https://doi.org/10.1037/a0014492

Imaizumi, S., & Tanno, Y. (2019). Intentional binding coincides with explicit sense of agency. *Consciousness and Cognition, 67*, 1–15. https://doi.org/10.1016/j.concog.2018.11.005

Ivanof, B. E., Terhune, D. B., Coyle, D., & Moore, J. W. (2022). Manipulations of Libet clock parameters affect intention timing awareness. *Scientific Reports, 12*(1). https://doi.org/10.1038/s41598-022-23513-1

Kliegl, R., Dambacher, M., Dimigen, O., Jacobs, A. M., & Sommer, W. (2012). Eye movements and brain electric potentials during reading. *Psychological Research Psychologische Forschung, 76*(2), 145–158. https://doi.org/10.1007/s00426-011-0376-x

Le, A. H., Burke, T., & Bayliss, A. P. (2026). Working with an online artificial partner enhances implicit and reduces explicit sense of agency. *Consciousness and Cognition, 137*, Article 103962. https://doi.org/10.1016/j.concog.2025.103962

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology, 49*(4), 764–766. https://doi.org/10.1016/j.jesp.2013.03.013

Libet, B. (2002). The timing of mental events: Libet's experimental findings and their implications. In *Consciousness and Cognition* (Vol. 11, Issue 2, pp. 291–299). Elsevier. https://doi.org/10.1006/ccog.2002.0568.

Majchrowicz, B., & Wierzchoń, M. (2018). Unexpected action outcomes produce enhanced temporal binding but diminished judgement of agency. *Consciousness and Cognition*. https://doi.org/10.1016/j.concog.2018.09.007

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

Pfister, R., Obhi, S. S., Rieger, M., & Wenke, D. (2014). Action and perception in social contexts: Intentional binding for social action effects. *Frontiers in human neuroscience, 8*, 667. https://doi.org/10.3389/fnhum.2014.00667

Posit team (2025). RStudio: Integrated Development Environment for R.Posit Software, PBC, Boston, MA. URL http://www.posit.co/.

R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Rezlescu, C., Danaila, I., Miron, A., & Amariei, C. (2020). More time for science: Using Testable to create and share behavioral experiments faster, recruit better participants, and engage students in hands-on research. *Progress in Brain Research, 253*, 243–262. https://doi.org/10.1016/bs.pbr.2020.06.005

Ruess, M., Thomaschke, R., Haering, C., Wenke, D., & Kiesel, A. (2018). Intentional binding of two effects. *Psychological Research Psychologische Forschung, 82*(6), 1102–1112. https://doi.org/10.1007/s00426-017-0892-4

Siebertz, M., & Jansen, P. (2022). Diverging implicit measurement of sense of agency using interval estimation and Libet clock. *Consciousness and Cognition, 99*. https://doi.org/10.1016/j.concog.2022.103287

Silver, C. A., Tatler, B. W., Chakravarthi, R., & Timmermans, B. (2025). The time course of Temporal Binding in social and nonsocial interactions. *Psychon Bull Rev* 32, 326–341 (2025). https://doi.org/10.3758/s13423-024-02540-1.

Stephenson, L. J., Edwards, S. G., Howard, E. E., & Bayliss, A. P. (2018). Eyes that bind us: Gaze leading induces an implicit sense of agency. *Cognition, 172*, 124–133. https://doi.org/10.1016/j.cognition.2017.12.011

Tanaka, T., Matsumoto, T., Hayashi, S., Takagi, S., & Kawabata, H. (2019). What makes action and outcome temporally close to each other: A systematic review and meta-analysis of temporal binding. *Timing & Time Perception, 7*(3), 189–218. https://doi.org/10.1163/22134468-20191150

Tonn, S., Pfister, R., Klaffehn, A. L., Weller, L., & Schwarz, K. A. (2021). Two faces of temporal binding: Action- and effect-binding are not correlated. *Consciousness and Cognition, 96*. https://doi.org/10.1016/j.concog.2021.103219

Uittenhove, K., Jeanneret, S., & Vergauwe, E. (2023). From lab-testing to web-testing in cognitive research: Who you test is more important than how you test. *Journal of Cognition, 6*(1), 13. https://doi.org/10.5334/joc.259

Vogel, D., Jording, M., Esser, C., Weiss, P. H., & Vogeley, K. (2021). Temporal binding is enhanced in social contexts. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-021-01928-7/Published

Vogel, D., Jording, M., Esser, C., Conrad, A., Weiss, P. H., & Vogeley, K. (2022). Temporal binding of social events less pronounced in individuals with Autism Spectrum Disorder. *Scientific Reports, 12*(1). https://doi.org/10.1038/s41598-022-19309-y

Wen, W., Yamashita, A., & Asama, H. (2015). The influence of action-outcome delay and arousal on sense of agency and the intentional binding effect. *Consciousness and Cognition, 36*, 87–95. https://doi.org/10.1016/j.concog.2015.06.004

Wenke, D., & Haggard, P. (2009). How voluntary actions modulate time perception. *Experimental Brain Research, 196*(3), 311–318. https://doi.org/10.1007/s00221-009-1848-8

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag.