Cognitive Psychology 151 (2024) 101662

Contents lists available at ScienceDirect

Cognitive Psychology

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

The fusion point of temporal binding: Promises and perils of multisensory accounts

Annika L. Klaffehn [a,*], Oliver Herbort [a], Roland Pfister [b,c]

[a] University of Würzburg, Germany
[b] Trier University, Germany
[c] Institute for Cognitive and Affective Neuroscience (ICAN), University of Trier, Germany

ARTICLE INFO

Keywords:
Temporal binding
Multisensory integration
Bayesian model

ABSTRACT

Performing an action to initiate a consequence in the environment triggers the perceptual illusion of temporal binding. This phenomenon entails that actions and following effects are perceived to occur closer in time than they do outside the action-effect relationship. Here we ask whether temporal binding can be explained in terms of multisensory integration, by assuming either multisensory fusion or partial integration of the two events. We gathered two datasets featuring a wide range of action-effect delays as a key factor influencing integration. We then tested the fit of a computational model for multisensory integration, the statistically optimal cue integration (SOCI) model. Indeed, qualitative aspects of the data on a group-level followed the principles of a multisensory account. By contrast, quantitative evidence from a comprehensive model evaluation indicated that temporal binding cannot be reduced to multisensory integration. Rather, multisensory integration should be seen as one of several component processes underlying temporal binding on an individual level.

1. Introduction

You press the button on a coffee machine and immediately see the little green light blink in readiness, hear the water heating up, smell the freshly ground beans and perhaps, already feel your mood lift. If you love coffee, that is. In any case, your cognitive system probably integrated all these related pieces of sensory information originating from very close temporal and spatial proximity into one coherent multimodal event (e.g., Debats et al., 2017). In retrospect it will likely be impossible to tell if you heard the water boiling first, or first smelled the coffee. Such integration of redundant and related sensory information is usually referred to as *multisensory integration* (e.g., Sanders et al., 2019). Although this mechanism entails the irretrievable loss of some detailed information, it crafts sparse and meaningful representations. This allows the human cognitive system to cope with the overwhelming wealth of everyday information and leads to a more stable and integrated perception.

Temporal and spatial integration of sensory information can be so strong that the single pieces of information are not processed individually but rather as a compound entity. Examples include visual-auditory integration as demonstrated by the well-known McGurk effect in which visual information changes heard syllables (McGurk & MacDonald, 1976) or the sound-induced flash illusion in which auditory information changes visual perception (Shams et al., 2000). One element of the coffee-making example stands out, however: the action of pressing the button on the coffee machine. From a logical point of view, this action must have occurred

* Corresponding author at: Röntgenring 11, 97070 Würzburg, Germany.
E-mail address: annika.klaffehn@uni-wuerzburg.de (A.L. Klaffehn).

https://doi.org/10.1016/j.cogpsych.2024.101662
Received 7 November 2022; Received in revised form 12 April 2024; Accepted 18 April 2024
Available online 20 May 2024

before any of the resulting effects. Nevertheless, actions and their effects are perceived as shifted towards each other in time, albeit classically not fused, as demonstrated by the temporal binding effect (originally termed intentional binding; Haggard et al., 2002).

Temporal binding refers to the perceptual illusion of subjectively shortened intervals between actions and subsequent effects. Here we ask whether temporal binding is a unique phenomenon tied to cause-effect sequences as has previously been proposed (Buehner, 2012), or whether it is better understood as representing incomplete multisensory integration (signal coupling; Ernst, 2006). Integrating temporal binding into the broad theoretical framework of multisensory integration would provide a parsimonious explanation of this intriguing and reliable illusion. Moreover, it would greatly inform the ongoing debate concerning its mechanistic and theoretical underpinnings (e.g., Klaffehn et al., 2021; Saad, Hemmer, & Musolino, 2022; Schwarz, & Weller, 2023; Wolpe et al., 2013). The following section, therefore, summarizes the main assumptions of multisensory integration accounts in the context of temporal binding.

1.1. Temporal binding as multisensory integration

What determines if multisensory signals become integrated? First, the signals need to fulfill the prerequisites of temporal and spatial proximity. Specifically the time (or delay) between individual signals is closely coupled to neural activation times (for more information see for example Stein & Stanford, 2008). Although integration might benefit from co-occurring peak activity, individual signals may be separated by up to 250 ms (Bedard & Barnett-Cowan, 2016; Noel et al., 2016) – the multisensory window of integration[1] (Chen et al., 2017) – and still be fused. Similarly, temporal binding of actions and their effects is constrained by the delay between the two (e.g., Haggard et al., 2002). Temporal action-effect binding, however, easily spans across delays of 400 ms (Ruess et al., 2017) or more (Humphreys & Buehner, 2009). Despite the differences in acceptable delay intervals, these findings can be reconciled. Note, that the multisensory window of integration refers to complete fusion of signals. This occurs when, coming back to the introductory example, we connect the sound of the coffee machine to the emerging smell of ground coffee resulting in a unified multisensory experience. Full fusion, however, is not necessarily expected regarding the action of pressing a button on the coffee machine: Logically, that action should have come first, and triggered the following (separable) event of the machine starting to work. It only follows that in classical findings on temporal binding actions and their effects are not usually perceived simultaneously. Cases of incomplete binding or signal coupling, akin to classical findings of temporal binding, have also been reported in the multisensory integration literature (Debats et al., 2017; Ernst, 2006, Wallace et al., 2004) and adhere to the rules laid out here.

A second principle governing multisensory integration is causal inference. According to it, multisensory illusions hinge on whether two sensory signals are perceived or expected to stem from a common event (Parise et al., 2012). The belief of a causal sequence is also at the heart of current theorizing on temporal action-effect binding. Not merely coincidence, but also causal beliefs shape the magnitude of the binding effect (Desantis et al., 2011, Fereday et al., 2019). Note that this cause may be an intentional action (following early assumptions; Haggard et al., 2002) and arguably, humans are especially attuned to detect causal connections related to their own actions and may even fabricate them to create illusions of control (Langer, 1975). However, the range of possible causes has been found to extend far beyond intention (Kirsch et al., 2019; Schwarz et al., 2019; but see Ruess et al., 2020; Weller et al., 2020). This highlights the importance of perceived causality rather than intention on the subjective temporal dislocation, and as such puts emphasis on the same principle as multisensory integration.

Finally, multisensory integration is also governed by the principle of inverse effectiveness. That is, how strongly a cue will be bound into a multisensory construct depends on features of this cue. Indeed, the perceptual strength of a sensory cue is inversely related to the perceptual enhancement it will receive if integrated with other cues (Holmes, 2009). Cues that are difficult to perceive when they are presented alone, as for example, the low rumbling of the coffee machine brewing water, will be integrated strongly into multisensory concepts. Expanding upon this principle, frameworks rooted in statistical optimality offer a more detailed insight into how different sensory cues may be integrated (Ernst & Banks, 2002). Such frameworks make use of maximum-likelihood estimates, where the variability (or uncertainty) of sensory cues is relevant when combining them. Cues that are difficult to perceive or pinpoint on their own carry less reliable information and multisensory constructs are therefore dominated by those cues, that are most easily perceived when presented alone.

Manipulating the perceptual strength of action and effects in a temporal binding setting led to results compatible with this statistically optimal implementation of the principle of inverse effectiveness. For example, making an effect less easily perceptible led to stronger binding of this effect towards the generating action and vice versa while leaving the absolute binding magnitude relatively untouched (Klaffehn et al., 2021; Wolpe et al., 2013). That is, an action performed on a force sensor (difficult to pinpoint in time) is attracted more strongly towards an ensuing effect tone than a standard keypress. Similarly, a tone that is difficult to hear over background noise is more strongly attracted towards an action than a tone that is easily perceptible. At the same time, recent research has also already questioned the efficacy of maximum-likelihood estimation to predict temporal binding. Tonn et al. (2021) showed that action binding (attraction of the action towards its effect) and effect binding (attraction of the effect towards its cause/the action) were not correlated across participants, suggesting that these two effects are surprisingly independent. A more comprehensive approach is therefore required to establish whether temporal binding can indeed be explained by the mechanisms of multisensory integration.

[1] Multisensory integration accounts have sometimes used the term of temporal binding window to characterize the window for complete fusion (e.g., McGovern et al., 2022; Ren et al., 2018; Venskus et al., 2021). These studies should not be confused with work on temporal binding of action-effect pairings despite the obvious terminological similarity.

1.2. Modeling temporal binding

Observing qualitative data patterns that conform to the predictions of a multisensory account is a promising first step towards integrating temporal binding into a multisensory framework (Klaffehn et al., 2021; Wolpe et al., 2013). On this basis formal models can be tested to establish whether temporal binding can indeed be attributed to multisensory integration.

An initial, encouraging attempt to provide a formal model for temporal binding was provided by Legaspi and Toyoizumi (2019) who effectively simulated group-level temporal binding data using a Bayesian model. In the current study we test predictions of a Bayesian model on novel temporal binding data. More specifically, we test whether temporal binding can be explained on an individual level by Statistically Optimal Cue Integration (SOCI) models (Ernst, 2006).

1.2.1. Statistically optimal cue integration (SOCI)

SOCI models are rooted in the conceptual underpinnings of Bayesian Decision Theory, a theoretical construct that furnishes a formal mechanism for comprehending how the cognitive system deals with inherent uncertainty (Ernst & Banks, 2002). That is, combining preexisting knowledge (i.e., the prior) with the sensory information we encounter enables us to make sense of our environment (i.e., the posterior). This approach has been successful in explaining human perception and performance across a wide range of tasks. SOCI models extend it to the context of multisensory integration, as in the example involving coffee-making. Here, the "prior" encompasses our assumptions and past experiences with similar situations. This prior is combined with sensory inputs related to making coffee, such as the smell, sound, and sight. This incoming information is naturally imprecise and uncertain. To improve perceptual precision and reduce redundancy our perceptual experience is shaped by combining the prior and all related sensory information, resulting in what we perceive as the "posterior", i.e., the experience of making coffee. Individual sensory cues affect this posterior based on their uncertainty. That is, less precise perceptual information carries less weight when generating the final perception from all sensory cues.

But how can an elusive concept such as perceptual uncertainty be quantified in mathematical terms? In the SOCI framework, it is approximated by isolating a sensory cue and repeatedly asking for perceptual judgements for that cue e.g., regarding location or timing. In case of an auditory cue, we could for example repeatedly ask, "When did you hear the tone?". These temporal judgments will naturally vary, most simply on a Gaussian distribution around the unbiased perception. A large variance of this distribution is indicative of less precise perception and therefore the variance of these judgements is used to quantify uncertainty about a cue's perception. This absolute uncertainty however reveals little about the weight, this cue will carry in the posterior. Instead, in the context of multisensory integration and its rule of inverse effectiveness the weight of a sensory cue in a final, unified perception is determined by its relative uncertainty compared to all other cues that are integrated as well. More simply, if two redundant cues are fused into one multisensory percept information from each cue is integrated into the final estimate according to their uncertainty as compared to the other cue(s). More weight is granted to less uncertain information. Using these weights, SOCI models compute the statistically optimal combination of all given information (e.g., regarding location or timing) and thereby predict human perception. Additionally, this combined perception should have a perceptual benefit over the single cues that constitute it. More precisely, uncertainty should be reduced through combination of information from all cues and perceptual judgements about the posterior should be more consistent than those derived from single cues.

In temporal binding paradigms researchers manipulate the perceived causal relationship between two events, typically a keypress and a tone. When participants infer a causal link, they often perceive the keypress (the action) and the tone (the effect) as occurring closer together in time. To allow predictions for these temporal shifts using the SOCI model, it is imperative to view both the action and the effect as distinct sensory signals, both of which convey information about a common temporal reference point. Within the framework of a SOCI model, we can then pinpoint this common timepoint in a statistically optimal manner. The original SOCI model, which we will refer to as the SOCI fusion model assumes a unified perception of the sensory cues at this common time point. As outlined in the coffee-making example, however, such full fusion is not likely to occur for action-effect sequences and, in fact, is not what is observed in common temporal binding setups.

SOCI models have been adjusted, though, to cover situations in which perfect fusion of cues is improbable. In this revised SOCI coupling model individual cues are attracted towards each other, but they are not fused at the shared reference point. Therefore, separate estimation points are predicted for each cue – in the context of temporal binding, for the action and the effect. The degree of attraction is determined by uncertainty about the stimuli's relationship, that is, uncertainty about whether or not fusion is appropriate. In mathematical terms, this is represented in form of a Bayesian prior. Effectively, a high uncertainty about the stimuli's relationship, such as under conditions of limited temporal or spatial proximity or conflicting top down expectations, will be reflected in a high variance of the Bayesian prior (Debats et al., 2017) leading to reduced predicted attraction. This incomplete fusion mechanism ostensibly serves to reduce the risk of erroneous coupling, but also comes with diminished benefits associated with multisensory integration.

1.2.2. The present study

We conducted two experiments to test whether the SOCI fusion model and the adaptation for situations without full fusion – the SOCI coupling model – provide a parsimonious multisensory explanation for the phenomenon of temporal binding. The precise mathematical building blocks used to construct and test these models are described in Section 2.3 and in Appendix A.

The experiments implemented situations where participants performed actions and triggered action-contingent auditory effects at varying action-effect delays. We further ensured that the delay conditions satisfied criteria for full perceptual fusion as well as sensory coupling, as evaluated in a pilot experiment (see Appendix B). Additionally, we implemented a condition with a sufficiently large delay to expect no temporal attraction between action and effect as well as one ambiguous situation, where interindividual differences in binding might be observed. For Experiment 1, we implemented action-effect delays of 50 ms (expected fusion), 150 ms (ambiguous situation), 400 ms (expected partial integration/ coupling) as per the pilot experiment, and 1000 ms (no binding expected, e.g., Haggard et al., 2002). In all conditions, we measured ratings of simultaneity as well as temporal binding via the Libet Clock procedure (see Fig. 1; e.g., Haggard, 2002), allowing us to compute the temporal binding illusion. We preregistered, amongst others, the following specific hypotheses:

- With very short action-effect delay (50 ms), participants subjectively report fusion of actions and effects. Mean estimates as well as individual standard deviations can be predicted by the SOCI fusion model.
- With medium action-effect delay (400 ms), participants do not report fusion, but actions and their effects are still temporally attracted, this temporal binding does not cover the full delay, however. Compared with the 50 ms delay condition, variances of actions and effect estimates deviate more strongly from the minimal variance predicted by the SOCI fusion model.
- With long action-effect delay (1000 ms), participants report no fusion, and actions and their effects are not temporally attracted. Variances of action and effect estimates are on baseline level.

With Experiment 2, we obtained an additional dataset comprising three distinct delay conditions (100 ms, 300 ms, 2000 ms) to reinforce and solidify our findings. The selection of the shorter two delay conditions aimed to observe the transition from fusion to coupling more clearly. The inclusion of the 2000 ms condition, by contrast, was prompted by an unpredicted observation of temporal binding even in the 1000 ms condition of Experiment 1. This longer delay condition was intended to guarantee a complete perceived dissociation between the executed action and its subsequent effect.

In the following analysis, we aimed to offer a thorough understanding of the data on a group level as well as on an individual level. In addition to assessing the precision of the SOCI model's predictions through the pre-registered comparison between observed estimates and model predictions, we enhanced our analysis with a likelihood ratio analysis, providing a more comprehensive measure of model fit. Prompted by the surprisingly mixed nature of these initial findings, we proceeded to test more general predictions related to multisensory integration, that rely less on the specifics of the model. To this end, we computed correlational markers to identify similarities between observed and predicted estimates and investigated whether temporal binding data can be predicted at the interindividual level by offering the same factors as those used in the SOCI models to an agnostic neural network model.

2. Methods

Preregistrations for Experiment 1 and 2, source code to adapt and run the experiment online, the full, anonymized dataset and the final analysis scripts are available at the Open Science Framework (osf.io/qgfns/).

2.1. Experiment 1

Fig. 1 gives an overview of the experimental design. The experiment featured a temporal binding setup in which participants provided time estimates for actions and action-contingent tone effects in different operant conditions. These conditions implemented a range of action-effect delays, intended to induce full action-effect fusion (50 ms), partial fusion (1 5 0), signal coupling (400 ms) and no fusion (1000 ms). The procedure additionally included baseline conditions with actions or tones appearing in isolation. These are

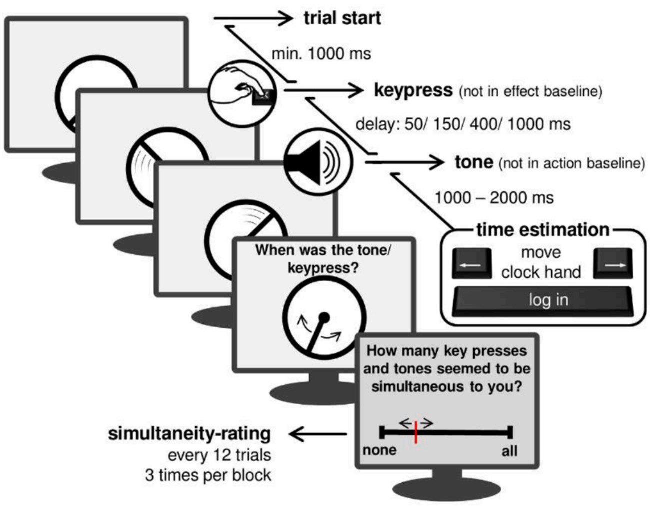

Fig. 1. Setup and procedure of Experiment 1. *Note.* Participants performed keypress actions and were presented with tones while watching a continuously rotating clock hand on screen. Actions or tones either appeared in isolation (baseline conditions) or the action consistently triggered a tone effect after a fixed action-effect interval (operant conditions). Participants then indicated the timing of either the keypress or the tone (blocked) by moving the clock hand to the position it had been at the time of the respective event. Action-effect delays were manipulated between experimental blocks, and participants rated at regular intervals how many of the last trials appeared to feature action and tone simultaneously.

necessary to compute unbiased action- and effect binding values as well as provide the markers for signal certainty employed in the SOCI model.

2.1.1. Participants

Desired sample size was set to 50 participants. This sample size offers a power of 1-β = 0.80 for effect sizes of $d_z = 0.40$ in paired *t*-tests, so that the experiment came with high power for common binding effects (e.g., $d_z = 0.50$ and 0.57 for action binding in free- and forced-choice actions and $d_z = 0.98$ and 1.03 for effect binding in Schwarz et al., 2019). The sample size further came with a power of 1-β = 0.80 for detecting moderate correlations of $r \geq 0.34$ between subjectively perceived fusion and actual temporal binding in a one-tailed test (computed with the *pwr* package in R).

Participants were recruited and compensated via the online recruitment platform Prolific (prolific.co). One dataset was rejected because it failed to meet the data quality criteria specified for participants in Prolific's experimental description (it came with completely random temporal estimates) and automatically replaced by another participant. Two more participants were excluded from the dataset, as they met exclusion criteria from the preregistration, but they were compensated and not replaced. Both failed to enter a temporal estimate (M + 2.5 SD) too many times. The final group of 48 participants reported a mean age of 27.5 years (±SD = 8.0; 18–54) and were mainly residents of the UK (12), Poland (7) and Italy (6) (detailed breakdown of nationality and country of residence in Appendix C). Participants self-identified as 19 female, 29 male and none as diverse and reported being 40 right-handed, 6 left-handed and 2 ambidextrous.

2.1.2. Stimulus material

The experiment was programmed with the JavaScript toolbox JsPsych (de Leeuw, 2015) and the additional psychophysics plugin (Kuroki, 2021) as an online experiment to be run in participants' internet browsers at home. It was provided on the free mindprobe server (mindprobe.eu) as sponsored by the European Society for Cognitive Psychology (ESCoP).

Each trial featured a Libet Clock consisting of a plain circle with a thick black outline and a continuously rotating clock hand extending from the circle center to the circle outline (see Fig. 1). The clock hand made a clockwise turn every 2500 ms. The clock face did not include any position markers. The size of the clock was determined by the screen of the participants' personal setup, and covered 400 pixels of the screen in diameter. The setup was further restricted to computers with physical keyboards (no cellphones or tablets). The experiment featured a 50 ms long 450 Hz beep tone which was used as action effect. To ensure participants had their volume sufficiently high, they had to complete an auditory detection task at three instances before and between experimental blocks where they responded as quickly as possible to the same tone, but with the volume lowered by the program. This task was only feasible with the volume turned up sufficiently and only after solving the task they could continue with the experiment. The auditory detection task had no experimental value but was a means to increase task adherence. The source file for the program as well as the auditory stimulus material can be accessed at osf.io/qgfns/.

2.1.3. Procedure

At the start of the session, the browser was automatically set to full screen. However, we did not track whether participants kept it on full screen during the whole experiment. After agreeing to the data protection and informed consent form, participants were asked to voluntarily enter their demographic data and the auditory detection task (see 2.1.2 Stimulus Material) was completed once to exclude all participants without working sound. Only participants who were able to detect the tones in this detection task were able to continue the experiment. Afterwards participants were introduced to the trial procedure briefly in two training trials. They were given the opportunity to repeat this introductory part if they wished, and otherwise they started with experimental trials.

A Libet Clock was presented from the start of each trial and its hand continuously rotated in clockwise direction. In all conditions except for the effect baseline condition, participants were to press the 'X' key of their keyboard with their left index finger. They could freely decide on the timepoint of this action as long as they waited at least one second after trial onset. Pressing the key too early triggered immediate error feedback and the trial was repeated. The keypress did not produce any visible changes in the display in any condition. In operant conditions however, the effect tone was played after a fixed action-effect delay of 50, 150, 400, or 1000 ms (varied blockwise). Afterwards the Libet Clock continued to rotate for a random duration of 1000 to 2000 ms. The clock hand then jumped to a random place on the clock and stopped moving. At this point, participants could move the clock hand by pressing the left and right arrow keys on their keyboard with their right hand. Participants were either asked to move the hand to where it had been when the key was pressed or when the tone was played. The event they were asked to estimate was blocked. Each delay condition comprised two blocks, each containing 36 trials: one for action estimates and one for effect estimates. Three times within these blocks (after 12 trials), participants were instructed to assess how many of keypresses and tones they had perceived as simultaneous since the last simultaneity rating or the beginning of the block. They provided ratings on a continuous rating scale ranging from 0 % to 100 %, with anchors labeled "none" and "all". Additionally, participants completed two baseline blocks (also consisting of 36 trials each): one where only keypresses were performed without hearing any tone (action baseline), and another where participants were instructed not to press any keys while tones were presented automatically (effect baseline). In both baseline conditions, participants were solely prompted to reproduce the clock hand position corresponding to the present event.

2.2. Experiment 2

Stimulus material and procedure were identical to Experiment 1 apart from the action-effect delays, which were set to 100, 300 and 2000 ms. The 2000 ms condition was implemented to ensure a condition without temporal binding. The 100 and 300 ms action-effect delay conditions were added to provide a more fine-grained picture of the transition from fusion to coupling.

2.2.1. Participants

In the second experiment, we did not plan to conduct any correlational analyses and therefore, opted for a sample size of 30 participants. Participants were recruited as in Experiment 1, and this time two datasets were rejected and automatically replaced by Prolific due to random ratings and non-compliance with instructions. One additional participant was excluded because they reported to have had trouble with the keyboard at the start of the experiment and indeed generated irregular data in the first blocks. They also met exclusion criteria from the preregistration and were not replaced. The final sample consisted of 29 participants aged 26.6 (±SD = 6.4; 18–41) years and residing mainly in South Africa (11), according to self-reported data. A detailed breakdown of nationality and country of residence is provided in Appendix C. Of this sample 14 participants self-identified as female and the rest as male and 23 reported themselves to be right-handed, 4 left-handed and 2 ambidextrous.

Due to the analysis methods we decided on *a posteriori* we still report exploratory correlations with this data. The final sample size of 29 participants would only detect large correlations of $r \geq 0.49$ with a power of 0.80 in a two-tailed test. The sample size should be sufficient for detecting temporal binding effects, as in Experiment 1 those ranged between $d_z = 0.55$ (1000 ms) and $d_z = 1.99$ (150 ms). The sample of Experiment 2 had a power of $1\text{-}\beta = 0.82$ to $1\text{-}\beta > 0.99$ to detect these effects in two-tailed tests.

2.3. Data preprocessing

The raw data consists of participants' estimates of actions and effects on the Libet Clock, as well as the logged true event timings. These estimates were recorded in milliseconds, with the clock hand's top position serving as the reference point (0–2500 ms). As a first step in preprocessing the data, we calculated estimation errors as the shortest temporal distance between the actual and estimated Libet Clock positions during cue presentation (clock- or counterclockwise). After excluding outlier trials falling beyond ± 2.5 SDs around a participant's condition mean, we aggregated these estimation errors across trials to estimate mean estimation errors (μ) and variances of estimation errors (σ^2) for each condition and each participant individually.

To assess temporal binding, we compared estimation errors in the baseline conditions ($\mu_{\text{bl\_action}}$, $\mu_{\text{bl\_effect}}$), where actions and tones were presented in isolation, to estimation errors in the operant condition ($\mu_{\text{op\_action}}$, $\mu_{\text{op\_effect}}$), where actions were followed by tones as effects. We computed action binding (i.e., the shift of the action towards the effect, $\mu_{\text{op\_action}} - \mu_{\text{bl\_action}}$, positive values indicate binding) and effect binding (i.e., shift of the effect towards the action $\mu_{\text{op\_effect}} - \mu_{\text{bl\_effect}}$, negative values indicate binding) separately. Finally, effect binding was subtracted from action binding to calculate overall temporal binding.

To gauge how well SOCI models fit temporal binding data, we computed estimation errors for the operant conditions as they would be predicted by two variations of the SOCI model: one assuming complete fusion of multisensory signals and the other assuming partial fusion (also signal coupling). In this context, we consider the action and the effect of the temporal binding setup as individual sensory signals, both conveying information about a common temporal reference point. A SOCI fusion model applied to this data posits that both action and effect are fused precisely at this common temporal point. We therefore computed a single statistically optimal estimate of the combined perceived time point for both action and effect (μ_{SO}). On the other hand, the coupling model suggests that the perceived timing of actions and effects should only be shifted towards this point by a certain degree. In this case, we calculate two distinct estimation errors ($\mu_{\text{SO\_action}}$, $\mu_{\text{SO\_effect}}$). Similarly, we can calculate the perceptual benefit through multisensory integration as the minimal variance the estimates in the operant condition should achieve if information were combined optimally. This minimal variance is again one common parameter for the SOCI fusion model (σ^2_{min}) whereas there are distinct parameters for the coupling model ($\sigma^2_{\text{min\_action}}$, $\sigma^2_{\text{min\_effect}}$).

The computation of perceived timing and minimal variance relies on the observed parameters in the baseline conditions of the temporal binding setup ($\mu_{\text{bl\_action}}$, $\mu_{\text{bl\_effect}}$, $\sigma^2_{\text{bl\_action}}$, $\sigma^2_{\text{bl\_effect}}$). As mentioned above, SOCI models estimate the uncertainty that is associated with the perception of a cue by isolating this sensory signal and calculating the variance across multiple perceptual judgments. The baseline conditions of our experiment, where keypresses and tones were assessed independently, served as suitable situations to estimate cue uncertainty ($\sigma^2_{\text{bl\_action}}$, $\sigma^2_{\text{bl\_effect}}$) and additionally provided unbiased temporal judgements of the keypress ($\mu_{\text{bl\_action}}$) and the tone ($\mu_{\text{bl\_effect}}$). Based on these absolute uncertainties ($\sigma^2_{\text{bl\_action}}$, $\sigma^2_{\text{bl\_effect}}$), we computed the relative uncertainty of the action and the effect as predicted by the SOCI fusion model for each participant individually as the weight (w), each of these cues would carry in the combined percept, $w_{cue} = \frac{\sigma^2_{\text{bl\_othercue}}}{\sigma^2_{\text{bl\_cue}} + \sigma^2_{\text{bl\_othercue}}}$. The weight of a cue (action or effect) increased, if the uncertainty (i.e., variance of judgements in the baseline condition) of all other cues was high as compared to the cue itself. As the action and the effect were the only two signals entered into the fusion model, the sum of their two weights was always 1. We computed the statistically optimal estimate (μ_{SO}) of a fused action-effect perception as the sum of the weighted baseline estimates ($\mu_{\text{bl\_action}}$, $\mu_{\text{bl\_effect}}$) of action and effect, $\mu_{\text{SO}} = (w_{\text{action}} * \mu_{\text{bl\_action}}) + \left(w_{\text{effect}} * \mu_{\text{bl\_effect}}\right)$. Finally, we computed the minimal variance (σ^2_{min}) this combined perception could achieve in the case of optimal maximum-likelihood estimation as $\sigma^2_{\text{min}} = \frac{\sigma^2_{\text{bl\_action}} * \sigma^2_{\text{bl\_effect}}}{\sigma^2_{\text{bl\_action}} + \sigma^2_{\text{bl\_effect}}}$ (Ernst & Banks, 2002, see also the Cramér-Rao bound; Cramér, 1946, Zachariah & Stoica, 2015).

The SOCI fusion model predicts one point in time, where action and effect are expected to be perceived simultaneously. The SOCI coupling model, however, merely posits attraction towards this timepoint and therefore calculates separate statistically optimal estimates for the action and the effect ($\mu_{\text{SO\_action}}$, $\mu_{\text{SO\_effect}}$). To account for such an incomplete cue integration (sensory coupling), a Bayesian coupling prior is included into the computation. This coupling prior represents assumptions made about the relationship of action and effect. As we mentioned before, participants are unlikely to expect perceptual fusion in action-effect situations as they would in other multisensory situations. This should be especially true when actions and effects are separated by longer delays. The coupling prior essentially represents the strength of the assumption of perceptual fusion. For example, if perceptual fusion in a situation is rather unexpected, uncertainty of the prior (σ^2_{prior}) is high and participants are unlikely to combine the action and the effect into a unified percept. However, the coupling prior cannot be directly measured or estimated from the baseline conditions. Still, we know how it should affect operant temporal binding conditions. For instance, if we observe only small temporal binding in a condition, possibly due to a large temporal delay between the action and the effect, this can only be represented in the SOCI coupling model as a variation in the coupling prior. Thus, we use the mean measured coupling strength in each situation across participants (i.e., proportion of the actual action-effect delay covered by temporal binding $\lambda_{\mathrm{mean\_pp}}$, see Debats et al., 2017), to approximate the uncertainty of the coupling prior as $\sigma^2_{\mathrm{prior}} = \frac{(1-\lambda_{\mathrm{mean\_pp}})}{\lambda_{\mathrm{mean\_pp}}} * (\sigma^2_{\mathrm{bl\_action}} + \sigma^2_{\mathrm{bl\_effect}})$ (see Appendix A). The coupling prior can then effectively be used to adjust the cue weights (w_{action}, w_{effect}) by the uncertainty of the coupling prior as $w_{cue} = \frac{\sigma^2_{\mathrm{bl\_othercue}}}{\sigma^2_{\mathrm{bl\_cue}} + \sigma^2_{\mathrm{bl\_othercue}} + \sigma^2_{\mathrm{prior}}}$. A large variance of the coupling prior consequently leads to reduced weights of the cues, resulting in less predicted binding. Based on these adjusted weights, statistically optimal estimation errors ($\mu_{\mathrm{SO\_action}}, \mu_{\mathrm{SO\_effect}}$) and minimal variances ($\sigma^2_{\mathrm{min\_action}}$, $\sigma^2_{\mathrm{min\_effect}}$) can be predicted for actions and effects individually and for each condition separately. See Appendix A for these computations as well as a summary of all computations used for calculating predicted estimation errors and minimal variances.

3. Analyses and results

We report exploratory analyses in addition to those preregistered. For simplicity, all tests are reported two-tailed. Effect sizes are reported as d = Cohen's $d_z = |t/\sqrt{n}|$. In case significant similarities were predicted, Bayes Factors are given as BF_{01}, with a larger Bayes Factor signifying stronger evidence for equality of two data sets.

3.1. Temporal binding

Analysis. In each condition, we assessed the presence of temporal binding. That is, we examined whether keypresses (actions) and tones (effects) were perceived as closer in time when the keypress triggered the tone as compared to a baseline condition where keypress and tone were estimated in isolation. As preregistered, we compared temporal binding in multiple conditions of Experiment 1. Specifically, we tested whether temporal binding in a condition where we expected fusion (50 ms delay) would cover more of the actual action effect delay than in a standard temporal binding condition with clear perceptual separation (400 ms condition). Additionally, we tested whether temporal binding would decrease in a condition considered to promote separate, independent perceptual representations (1000 ms condition).

Additionally, we conducted a comparative analysis of explicit reports of perceived fusion, using predefined thresholds of 75 % (indicating the experience of fusion) and 25 % (indicating absence of fusion). These thresholds, established prior to data collection, were selected for pragmatic purposes due to a lack of guiding empirical evidence. These thresholds may serve as useful benchmarks for future research in this area.

Results. Fig. 2 summarizes the empirical results. In all conditions but the 2000 ms delay condition, participants showed significant temporal binding, whereas in the 2000 ms delay condition, we observed significant repulsion. The ratio of action-binding to effect-binding consistently increased with increasing action-effect interval. With an increase from 50 to 400 ms action-effect delay, the overall temporal binding increased from 95 ms to 150 ms, $t(47) = 2.07$, $p = .044$, $d = 0.30$, but the proportion of the actual action-effect delay covered by binding decreased, $t(47) = 5.85$, $p < .001$, $d = 0.84$ (see Fig. 2, Boxplots see Appendix D, Fig. D1). Despite the large delay of 1000 ms, temporal binding was not reduced in this condition as compared to the 400 ms condition, $t(47) = 0.38$, $p = .708$, $d = 0.05$.

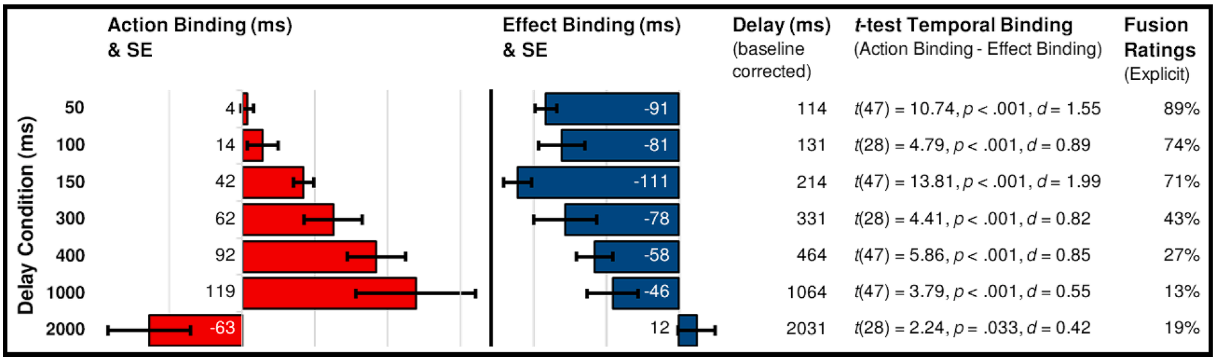

Fig. 2. Binding and explicit fusion ratings in all conditions. *Note.* Action and effect binding (operant – baseline) in all conditions of Experiment 1 and 2. Error bars show standard errors of the mean (SE). They are accompanied by the delay between baseline estimates of keypresses and tones, t-tests for the combined binding effect (action binding − effect binding) and the mean percentage of trials in which action and effect had been perceived to occur simultaneously (Fusion Ratings).

Explicit fusion ratings decreased continuously with increasing delay. In Experiment 1, participants reported the most fusion with a delay of 50 ms (89 %) which was significantly more than in the 150 ms condition (74 %, $t(47) = 5.55$, $p < .001$, $d = 0.80$) which was in turn more than in the 400 ms condition, (26 %, $t(47) = 10.20$, $p < .001$, $d = 1.47$).

In the 50, 100 and 150 ms condition explicit ratings exceeded or were similar to the preset upper threshold of 75 % (50 ms: $t(47) = 6.31$, $p < .001$, $d = 0.91$; 100 ms: $t(28) = 0.30$, $p = .764$, $d = 0.06$, BF = 6.66; 150 ms: $t(47) = 1.16$, $p = .252$, $d = 0.17$ BF = 4.61). Note though, that action and effect binding combined never spanned the complete, baseline corrected delay interval. Actions and effects were thus always judged at different points in time (action estimate vs. effect estimate 50 ms: $t(47) = 2.29$, $p = .026$, $d = 0.33$; 100 ms: $t(28) = 2.22$, $p = .034$, $d = 0.41$; 150 ms: $t(47) = 4.35$, $p < .001$, $d = 0.63$), providing no objective evidence for perceptual fusion. In the 300 ms condition, despite a delay surpassing common multisensory integration windows, participants judged more than 25 % of all trials as having been perceived simultaneous, $t(28) = 3.88$, $p < .001$, $d = 0.72$. Explicit fusion ratings in the 400 ms condition were similar to 25 %, $t(47) = 0.41$, $p = .684$, $d = 0.06$, BF = 8.16. Additionally, in the 1000 ms and 2000 ms delay conditions, participants explicit fusion ratings were similar or smaller than 25 %, 1000 ms: $t(47) = 3.83$, $p < .001$, $d = 0.55$, 2000 ms: $t(28) = 1.32$, $p = .198$, $d = 0.24$, BF = 3.07.

3.2. SOCI model fit

3.2.1. Log Likelihoods (−2LL)

Analysis. To determine in how far the temporal binding values (see Fig. 2) could be explained by SOCI models, we predicted estimates as well as minimal variances with a fusion and a coupling model based on statistically optimal cue integration for each participant individually. We then computed the likelihood of each of the observed ratings in the operant condition given the predicted distribution of estimates. The likelihoods of ratings in individual trials were aggregated into separate −2LL values for actions and effects for each participant and each delay condition (Lewandowsky and Farrell, 2011, see Appendix E for the computations). To assess the explanatory power of the SOCI coupling and fusion models, we created simpler fusion and coupling models as comparison models (simple models) with fixed, equal weights for actions and effects. The simple fusion model thus had fixed weights of 0.5, whereas the simple coupling model had equal fixed weights, that were each half of measured coupling strength. In the simple models we assumed variance of operant action and effect estimates to remain equal to the baseline variances. On the basis of these models, we also computed the likelihood of observed values in the operant conditions. None of the models included any free parameters. With *t*-tests we compared −2LLs of the simple models with those of the SOCI models for each condition separately.

Results. Fig. 3 shows the log likelihood results for the SOCI fusion and coupling models relative to their simple counterparts with fixed weights. The SOCI fusion model fit action estimates in the 50 ms condition numerically better than the simple fusion model, $t(47) = 1.97$, $p = .054$, $d = 0.28$, BF = 1.41, whereas the SOCI coupling model performed significantly better than its simple counterpart, $t(47) = 2.16$, $p = .036$, $d = 0.31$. The only other condition with a significant benefit of any SOCI model over the simple model were the action estimates in the 2000 ms delay condition of Experiment 2, $t(28) = 3.46$, $p = .002$, $d = 0.44$ (coupling model). Remarkably, for most effect estimation data, the simple model performed at least as well as the SOCI model.

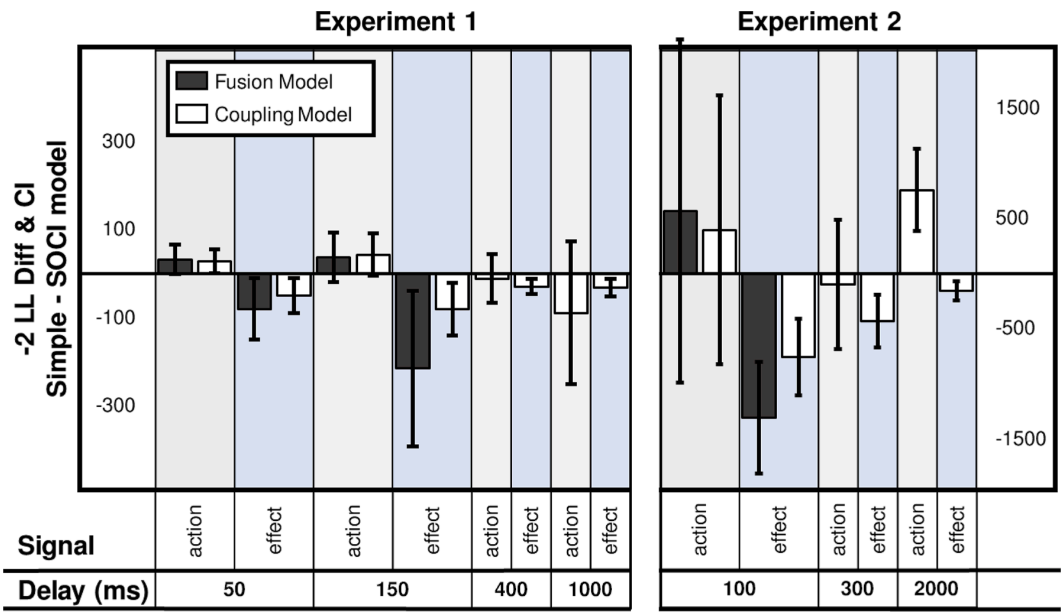

Fig. 3. Results of the Log Likelihood (LL) analysis. *Note.* We generated the theoretical distribution of estimates using SOCI fusion and SOCI coupling models. Subsequently, we aggregated −2LL values for participants' actual estimates and compared them with −2LL values for simple comparison models that assumed equal weight for action and effect as components of temporal binding. The figure illustrates the difference between these −2LL values (simple − SOCI model) for each condition, accompanied by 95% confidence intervals. Positive values indicate a higher likelihood for the SOCI model, while negative values favor the simple model. The data is categorized by delay condition, predicted signal (action/effect), and model assumptions utilized for predictions (fusion/coupling).

3.2.2. Exactness of predicted estimates

Analysis. Following the preregistered analyses, we tested whether observed operant action and effect estimates were numerically similar to or different from what would be predicted by SOCI fusion and coupling models with a Bayesian *t*-test. Bayes factors (BF_{01}) ≥ 3 were interpreted as evidence for similarity. The numerical accuracy of estimates for the SOCI fusion model was only assessed for the short delays (50 ms, 100 ms, 150 ms), as a precise fit is unlikely with longer delays.

Results. Fig. 4 compiles mean action and effect estimates as well as associated probability density functions for all conditions. Observed action estimates corresponded to predictions by the SOCI coupling model only in the long conditions of 1000 ms, $t(47) = 1.32$, $p = .193$, $d = 0.19$, BF = 3.82, and 2000 ms, $t(28) = 1.24$, $p = .224$, $d = 0.23$, BF = 3.34 (negative coupling prior). Observed and predicted effect estimates were similar for the SOCI fusion model for effect binding in the 100 ms condition, $t(28) = 0.98$, $p = .334$, $d = 0.18$, BF = 4.39. Predictions based on the SOCI coupling model were similar to observed effect estimates in the 100 ms condition, $t(28) = 0.11$, $p = .916$, $d = 0.02$, BF = 6.93, the 150 ms condition, $t(47) = 0.50$, $p = .621$, $d = 0.07$, BF = 7.85, and the 300 ms condition, $t(28) = 0.60$, $p = .555$, $d = 0.11$, BF = 5.86 (see Fig. 4). Additionally, more perceived fusion of actions and effects in the 150 ms condition was related to a smaller difference between observed effect estimates and those predicted with the SOCI fusion model, $r(46) = -.55$, $p < .001$. This was not the case for action parameters.

3.2.3. Exactness of predicted variances

Analysis. Multisensory integration assumes an improvement in perceptual precision through sensory integration. The SOCI models predict a minimal possible variance of the integrated percept(s). With Bayesian *t*-tests we tested whether observed variances were numerically similar to this prediction. Numerical accuracy of SOCI fusion models was again only assessed for short delays (50 ms, 100 ms, 150 ms).

Results. Observed variances of action estimates were numerically bigger than what was predicted by the SOCI fusion or coupling model across conditions (see Fig. 4). Predictions by the SOCI fusion model were never similar to observed variances. However,

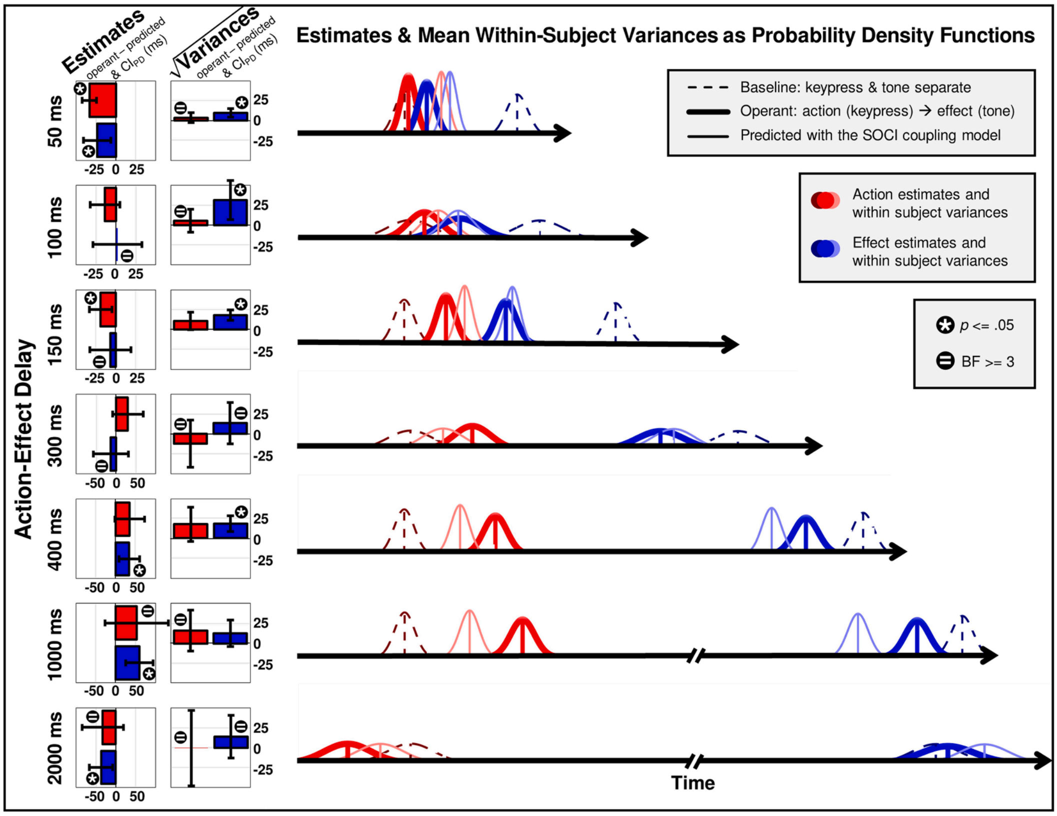

Fig. 4. Operant vs. predicted action and effect estimates and within subject SDs. *Note.* On the left hand side, temporal estimates and within-subject SDs as they were observed in the operant condition are contrasted with predictions from the SOCI coupling model. The respective difference (operant − predicted) is shown on bar graphs along with the 95 % CI of the paired differences. These are sometimes annotated with symbols demarking the results of a Bayesian *t*-test for this contrast with (*) for significant differences ($p <= .05$), and (=) for significant similarities between operant and predicted values (BF >= 3). To the right, these values are represented on the trials timeline, with vertical lines marking the time, estimates were given for actions (red) and effects (blue) in baseline trials (dashed line), operant trials (heavy line) and where estimates were predicted to be given in operant trials based on the SOCI coupling model (light line). The accompanying bell curves show within-subject variances as observed in baseline and operant trials or as predicted by the SOCI coupling model. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

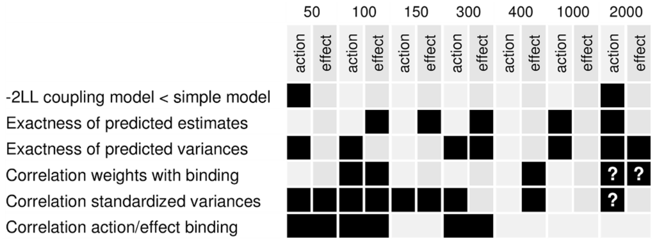

Fig. 5. Results in favor of the SOCI coupling model. *Note.* Summary of the results in all the analysis methods we applied. When a test favored the SOCI coupling model for a condition, the corresponding analysis-condition intersection is shaded in black. "?" = significant relationship, but expected direction is not clear.

observed variances of operant actions were similar to what was predicted by the SOCI coupling model in the 50 ms condition, $t(47) = 1.02$, $p = .315$, $d = 0.15$, BF = 5.37, the 100 ms condition, $t(28) = 0.75$, $p = .462$, $d = 0.14$, BF = 5.33, the 300 ms condition, $t(28) = 0.80$, $p = .432$, $d = 0.15$, BF = 5.13 as well as in the two long conditions of 1000 ms, $t(47) = 1.22$, $p = .230$, $d = 0.18$, BF = 4.33, and 2000 ms, $t(28) = 0.02$, $p = .982$, $d < 0.01$, BF = 6.97. Variances of operant effects were only similar to SOCI coupling predictions in the 300 ms condition, $t(28) = 1.02$, $p = .314$, $d = 0.19$, BF = 4.22, and in the 2000 ms condition, $t(28) = 1.01$, $p = .322$, $d = 0.19$, BF = 4.28 (see Fig. 4). Even though in the 150 ms condition, effect variances were different to those predicted, increased perceived fusion of actions and effects was also related to a smaller difference between observed effect variances and those predicted with the SOCI fusion model, $r(46) = -.53$, $p < .001$. Again, this was not true for action parameters.

3.3. Multisensory integration fit

Limited predictive power of the SOCI model does not necessarily indicate that multisensory integration principles in general fail to capture temporal binding. To gain a more comprehensive understanding of our data, we conducted supplementary tests that allowed for greater flexibility beyond the specific predictions of the SOCI model. Exact results are reported in appendices and are only summarized here. An overview of which tests in which conditions yielded favorable results for the applicability of the SOCI coupling model onto temporal binding can be seen in Fig. 5.

First, we examined the correlations between predicted estimates and their associated variances with their observed counterparts. This provided insights into whether the data exhibited a multisensory integration pattern, even in cases where exact predictions were not accurate. Estimates showed no clear picture in this analysis, whereas variances clearly correlated in the smaller delay conditions (50 ms, 100 ms, 150 ms, see Appendix F).

Additionally, we tested a more general assumption of multisensory integration, namely the negative association of action binding with effect binding. This association was present in the 50 ms, 100 ms and 300 ms condition (see Appendix F). Finally, to test whether multisensory integration factors could effectively predict estimates in a temporal binding setting when not constrained by a specific model, we employed an impartial neural network, and trained it using temporal binding data from four previously published datasets (Kirsch et al., 2019; Klaffehn et al., 2021; Ruess et al., 2017; Schwarz et al., 2019). We tasked this neural network with utilizing action- and effect binding values, along with their respective variability and the temporal delay between action and effect, to forecast estimates and variances within the present data. The model exhibited overall subpar performance in these predictions, particularly in the case of action estimates (see Appendix G).

4. Discussion

Multisensory integration has been discussed as a promising candidate mechanism to explain the perceptual illusion of temporal binding for over a decade (e.g., Wolpe et al., 2013). Motivated by empirical studies producing qualitative patterns that follow multisensory principles (e.g., Klaffehn et al., 2021) and modelling efforts on a group level (Legaspi & Toyoizumi, 2019), we used the computational approach of the multisensory SOCI model to test if temporal binding can indeed be viewed as a pure instance of multisensory integration.

Participants completed a temporal binding task, where they performed keypresses and thereby produced a tone. The tone was delayed by 50, 150, 400 or 1000 ms across different blocks of Experiment 1 and it was delayed by 100, 300 or 2000 ms across the blocks of Experiment 2. Participants were asked to estimate the timing of action or effect on a Libet Clock (also varied blockwise). They were further periodically asked to estimate how many actions and tones seemed to have been simultaneous over a series of trials. We used this dataset to test whether a computational model of multisensory integration, the SOCI model, was able to account for temporal binding on an interindividual level. To draw comprehensive conclusions, we tested the model for different action-effect delays, where actions and effects would be perceived as fused (50 ms, 100 ms), where they would be perceived to be shifted toward each other but appear clearly separate (300 ms, 400 ms), where no binding should be expected (1000 ms, 2000 ms) as well as one ambiguous situation, where we specifically expected interindividual differences regarding the extent to which fusion was experienced (150 ms).

4.1. Temporal binding and model fit at group level

Participants showed clear and stable temporal binding of actions and effects in both experiments. Binding was observed even when actions and effects were separated by a 1000 ms delay. Indeed, it was only absent with a 2000 ms delay, in which case actions and effects were repulsed. In the shortest delay conditions, participants reported high percentages of perceived action-effect fusion although no objective fusion was present in the estimates.[2]

Importantly, a clear trend emerged in that action binding decreased and effect binding increased as the action-effect delay lengthened (see Fig. 2). With longer delays, the time until the action rating could be given also increased. Simple memory decay therefore reasonably led to diminished certainty about action estimates in the longer delay conditions. At the same time, the effects were always followed by the rating phase with the same delay of 1000–2000 ms, resulting in constant certainty about the effect timing. In line with multisensory integration theories, the decline in action certainty should lead to a shift in the weight ratio between action and effect binding, tilting it away from the action and toward the effect. In longer delay conditions the perceived timing of the effect should thus carry significantly more weight, predicting the observed increase in action binding and decrease of effect binding. This finding would present substantial evidence for the impact of multisensory integration principles on a group level. Nevertheless, it is important to acknowledge that this interpretation of the data pattern is constrained by additional observations. For one, a straightforward consequence of this explanation for the shift in the action-to-effect binding ratio would entail an extreme change in the certainty of action estimates through the increased delay. Although action variances in the operant condition did increase with increasing delay, their change did not reach the extent predicted by the ratio shift. Secondly, we tested post-hoc whether different rating delays would have a substantial impact on rating certainty (i.e., variance of the estimates). This was not the case.[3] In fact, this anecdotal evidence raises a more fundamental question regarding whether baseline variances effectively capture trial-to-trial perceptual uncertainty.

The finding of stable temporal binding with an increasing portion of it being conveyed over the action binding element persisted even with an action-effect delay of 1000 ms. Findings of temporal binding with such a large delay had previously been confined to the interval estimation method (Humphreys & Buehner, 2009) as opposed to paradigms using the Libet Clock as employed in the present experiments (Haggard et al., 2002). In our case, the reduced layout of the Libet Clock combined with the broad range of action-effect delays might have contributed to the finding (but see Ivanof et al., 2021 for a limited impact of such design features on temporal binding).

In Experiment 2, we employed an even longer delay between actions and effects to ensure a condition without temporal binding. Indeed, in the 2000 ms delay condition, we found a significant repulsion of actions and effects. The action-effect delay may thus have been sufficiently large to decouple action and effect and lead to a perception that fosters separation, as has occasionally been speculated in the literature (e.g., Haggard et al., 2002). Notably, however, the aforementioned trend of the action to effect binding ratio across delay conditions persisted even within the temporal repulsion of the 2000 ms condition. This finding may therefore alternatively highlight the influence of spatial binding in Libet Clock paradigms. In fact, only in the 2000 ms condition, the clock hand has almost completed a full rotation by the time the effect appears. The closest spatial connection is therefore the inverse of the distance traversed during the delay. Although evidence from other paradigms, most notably interval estimates (Humphreys & Buehner, 2009), shows clear temporal binding without the possible influence of spatial binding, it's impact on in Libet Clock paradigms may be of great interest if confirmed.

4.2. Model fit at the individual level

We set out to investigate fit of the SOCI model, especially the coupling version, to our current data. Analyses of Log Likelihoods as a measure of how well different models captured the empirical data suggested the fit of the SOCI model to be suboptimal. Crucially, comparing SOCI fusion and coupling models to simple models that assumed equal weights yielded no advantage for the more complex model in almost all conditions (see Fig. 3). At the same time, the SOCI models were able to predict several specific estimates and their associated variances. This was demonstrated through paired *t*-tests that compared observed estimates and variances with their predicted counterparts. However, it is important to note that the instances where observations and predictions significantly aligned did not exhibit a discernible pattern. Instead, these instances were scattered across various delay conditions and encompassed both action and effect binding.

We see three possible interpretations of these volatile results. Firstly, we could of course assume that multisensory integration does not play a role in temporal binding paradigms. This seems unlikely for several reasons, including existing findings that actions and effects may be bound in a multisensory fashion (e.g., Debats et al., 2017). Additionally, a contribution of multisensory integration to temporal binding has been established in previous work (Klaffehn et al., 2021, Legaspi & Toyoizumi, 2019, Wolpe et al., 2013) and indeed, when examined at the group level, the relationship of action and effect binding in the current dataset also indicates the influence of multisensory integration principles. So alternatively, our current findings may suggest that temporal binding is in part

[2] As actions and effects were not estimated simultaneously but in separate blocks, this may not indicate actual separation of the two cues but alternative explanations are possible, such as systematic judgment errors.

[3] Certainty was almost identical, irrespective of whether the baseline condition was given with a 1000–1400 ms delay ($\mu_{Exp1} = 1204$ ms; $\mu_{Exp2} = 1196$ ms) or with a 1600–2000 ms delay ($\mu_{Exp1} = 1799$ ms; $\mu_{Exp2} = 1813$ ms); Exp1: $\sigma_{short} = 74.2$ ms vs. $\sigma_{long} = 71.3$ ms; Exp2: $\sigma_{short} = 92.1$ ms vs. $\sigma_{long} = 93.4$ ms.

influenced by or akin to multisensory integration, and that this part can indeed be captured by the SOCI model. Yet, there are more underlying parameters that obscure this relationship. These may especially include person-specific parameters as illustrated by the good performance of a Bayesian model at the group level (Legaspi & Toyoizumi, 2019). Finally, the differences between the SOCI model and the empirical data, especially the poor performance as compared to a simple model in terms of Log Likelihoods might highlight the inaptness of the model rather than multisensory integration per se.

To further investigate the fit of multisensory integration rather than the specific SOCI model, we conducted increasingly flexible analyses on the data. Still mostly keeping in line with the SOCI model, we correlated standardized estimates (i.e., observed weights) and variances with their predicted counterparts. Notably, standardized estimates rarely correlated significantly with predictions generated by the SOCI coupling model. This additional insight also qualifies the preceding conclusions regarding estimation precision. In cases where estimates and predictions exhibited no correlation, the absolute precision of the prediction may hold less significance. Conversely, observed variances consistently displayed correlations with predicted variances, particularly within the shorter delay conditions (50–150 ms).

Based on a more general assumption of multisensory integration, we additionally tested whether action binding and effect binding would be negatively correlated. Indeed, in three shorter delay conditions (50, 100 and 300), action and effect binding were related in the predicted way, giving further indication that multisensory integration principles are relevant in this setting. Yet, the resulting pattern of all tests reported above as shown in Fig. 5 attests to striking limitations regarding specific as well as more general model fit to our data. Finally, breaking completely away from the confines of the SOCI model and only taking the relevant predictors to an agnostic neural network, this also revealed a suboptimal capacity to predict the observed action and effect parameters. This highlights that it may not simply be the combination of data in the SOCI model which fails to provide good predictions, but that these predictors employed for multisensory integration are insufficient to reliably compute operant parameters, in particular action estimates. Additionally, this finding may indicate that on an intraindividual level the mean baseline variance is insufficient to predict perceived certainty of a cue.

4.3. Limitations

While the SOCI models served as a promising initial framework, it is essential to acknowledge their inherent limitations in capturing the complexity of multisensory integration and future research should explore alternative models. Furthermore, there may be potential for an improved intraindividual model fit by increasing the number of trials, warranting further investigation. Moreover, we cannot exclude that the imperfectly matched action baseline conditions for extended action-effect delays impacted model fit in these conditions – an important design factor that should be considered in further applications of this paradigm.

5. Conclusions and outlook

Temporal binding assessed on the group level complies to qualitative predictions of multisensory integration models as evidenced by data patterns across different delay conditions. However, comparing empirical data of individuals to mathematical predictions from multisensory integration models (Statistically Optimal Cue Integration) highlights remarkable inconsistencies. In fact, a neural network trained to predict binding measures based on these variables did not yield substantial success. It is reasonable to assume that the quantification of perceptual certainty via baseline variances, a core aspect of SOCI models, fails to completely capture personal perceptual certainty in this situation.

The notion that temporal binding data is marked by large interindividual differences has been proposed before (Lush et al., 2019, Saad, Hemmer, & Musolino, 2022). Indeed, the success in showing multisensory integration principles in temporal binding data on a group level (Klaffehn et al., 2021; Legaspi & Toyoizumi, 2019; Wolpe et al., 2013) in contrast to inconclusive data on an individual level, as in the current manuscript and by Tonn et al., 2021 greatly support this assumption. It will doubtless be worthwhile to consider more personal, state-dependent factors as well as situational influences when attempting to model temporal binding on an individual level.

The overall limited explanatory strength of multisensory integration accounts for temporal binding data underscores the need to explore additional contributing factors as well as alternative theoretical frameworks. One candidate model, proposed by Saad et al. (2022), offers a fundamentally distinct mechanistic approach to temporal binding. The authors show that temporal binding measured by interval estimations (reproducing or judging the time between an action and an effect) can be simulated with a Bayesian rational memory model. Although this model capitalizes on memorizing different time intervals and therefore the transferability to Libet Clock paradigms is not straightforward, memory effects offer an intriguing new approach to temporal binding findings. In a related vein, Schwarz & Weller (2023) highlighted the impact of attention on time perception also possibly influencing temporal binding results profoundly.

As the underlying mechanistic explanation for the consistent and readily replicable phenomenon of temporal binding remains elusive, the path forward may indeed involve the synthesis of more complex models as a combination of multiple existing theoretical frameworks.

Statement: During the revision of this work the authors used Chat GPT in order to generate alternative phrasings for single sentences. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

ALP received funding by ev. Studienwerk Villigst. RP is funded by the German Research Foundation (PF853/10-1, PF853/11-1).

CRediT authorship contribution statement

Annika L. Klaffehn: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Oliver Herbort:** Writing – review & editing, Supervision, Formal analysis, Data curation. **Roland Pfister:** Writing – review & editing, Supervision, Methodology, Data curation.

Data availability

Data can be accessed here: https://osf.io/qgfns/.

Appendix A. Data processing

Baseline Action:

Participants press a key without triggering a tone effect. They then estimate the time of their keypress on the Libet clock by bringing the clock hand into the position they believe it had been at, when they pressed the key. For each estimate, we computed the perceptual shift relative to the actual onset of the keypress, using the shortest distance in ms when going either clockwise or counter-clockwise. Negative numbers indicate estimates before the actual onset, positive numbers indicate estimates after the actual onset. For each participant, we compute mean and variance of the distribution across trials and use these as the parameters $\mu_{\text{bl\_action}}$ and $\sigma^2{}_{\text{bl\_action}}$.

Baseline Effect:

Participants estimate the time of a tone stimulus without pressing a button. For each estimate, we computed the perceptual shift relative to the actual onset of the tone. Negative numbers indicate estimates before the actual onset, positive numbers indicate estimates after the actual onset. For each participant, we compute mean and variance of the distribution across trials and use these as the parameters $\mu_{\text{bl\_effect}}$ and $\sigma^2{}_{\text{bl\_effect}}$.

Operant Action:

Participants press a key in order to trigger a tone. This tone may be delayed by various delay intervals. They then estimate the onset of the keypress (irrespective of the tone), and we computed the perceptual shift relative to the actual keypress. For each participant and each delay condition, we compute mean and variance of the distribution across trials and use these as the parameters $\mu_{\text{op\_action}}$ and $\sigma^2{}_{\text{op\_action}}$.

Operant Effect:

Participants press a key in order to trigger a tone. This tone may be delayed by various delay intervals. They then estimate the onset of the tone (irrespective of the keypress), and we computed the perceptual shift relative to the actual tone. For each participant and each delay condition, we compute mean and variance of the distribution across trials and use these as the parameters $\mu_{\text{op\_effect}}$ and $\sigma^2{}_{\text{op\_effect}}$.

Action binding ($\mu_{\text{op\_action}} - \mu_{\text{bl\_action}}$):

Perceived shift of the action towards the effect. Positive values indicate a shift forward in time and therefore towards the effect.

Effect binding ($\mu_{\text{op\_effect}} - \mu_{\text{bl\_effect}}$):

Perceived shift of the effect towards the action. Negative values indicate a shift backward in time and therefore towards the action.

Temporal binding (Action binding – Effect binding):

Integration of action binding and effect binding into one measure. Higher values indicate a larger shift of actions and effects towards each other.

Model computations

SOCI fusion model

In applying the SOCI fusion model to temporal binding data, we assume that actions and effects are sensory cues that each provide information about a common point in time. This temporal point can be estimated by integrating information from both cues and weighting them based on the uncertainty with which each sensory cue is perceived. This uncertainty can be approximated from the data with the parameters $\sigma^2{}_{\text{bl\_action}}$ and $\sigma^2{}_{\text{bl\_effect}}$. In a simple scenario involving just two cues, weights are based on a comparison of the uncertainty associated with the other cue and its influence on the final perception in relation to the first cue. When the uncertainty of the other cue is high, it leads to a higher weight assigned to the first cue.

Weights (w)

$$w_{\text{action}} = \frac{\sigma^2{}_{\text{bl\_effect}}}{\sigma^2{}_{\text{bl\_action}} + \sigma^2{}_{\text{bl\_effect}}} \tag{1}$$

$$w_{\text{effect}} = \frac{\sigma^2{}_{\text{bl\_action}}}{\sigma^2{}_{\text{bl\_action}} + \sigma^2{}_{\text{bl\_effect}}} \tag{2}$$

$$w_{\text{action}} + w_{\text{effect}} = 1 \tag{3}$$

From these weights, as well as unbiased action and effect estimates (μ_{bl_action}, μ_{bl_effect}) the SOCI fusion model can compute the statistically optimal estimate (μ_{SO}). This calculation involves aligning the baseline estimates of the action (μ_{bl_action}) and the effect $\left(\mu_{bl\_effect}\right)$ to a common timeframe, which is dependent on the action-effect delay in the specific operant condition for which the statistically optimal estimate is being determined. For example, if the action and the effect are separated by 300 ms in the operant condition, 300 ms should be added to the mean estimation error of the effect baseline (μ_{bl_effect})

$$\mu_{\mathrm{SO}} = (w_{\mathrm{action}} \times \mu_{\mathrm{bl\_action}}) + \left(w_{\mathrm{effect}} {}^* (\mu_{\mathrm{bl\_effect}} + \mathrm{delay})\right) \tag{4}$$

Statistically optimal estimates therefore vary depending on the specific action-effect delay being predicted.

Ultimately, a statistically optimal estimate should be marked by higher precision compared to the individual sensory cues from which it is derived. Optimal estimates should approach the Cramér-Rao bound. In our case, the minimal variance of the combined perception ($\sigma^2{}_{min}$) can be computed as

$$\sigma^2{}_{\mathrm{min}} = \frac{\sigma^2{}_{\mathrm{bl\_action}} \times \sigma^2{}_{\mathrm{bl\_effect}}}{\sigma^2{}_{\mathrm{bl\_action}} + \sigma^2{}_{\mathrm{bl\_effect}}} \tag{5}$$

SOCI coupling model

The SOCI coupling model introduces uncertainty about the fusion of actions and effects. This uncertainty should be mainly based on the delay condition with larger delays leading to more uncertainty about the relationship of actions and effects. Participants' behavior provides an approximation of this uncertainty. Given that the SOCI coupling model posits reduced temporal attraction as a consequence of heightened uncertainty about the relationship of actions and effects, this uncertainty can be inversely estimated from the temporal binding (Action binding – Effect binding) observed. To obtain an impartial estimate of temporal attraction in the operant condition, it is essential to compare the time between operant estimates to the theoretical time between baseline estimates, the baseline action effect delay, assuming they were provided in the same sequence but without any bias,

$$\mathrm{Baseline\,Action\,Effect\,Delay} = \mu_{\mathrm{bl\_effect}} + \mathrm{delay} - \mu_{\mathrm{bl\_action}} \tag{6}$$

Delay in this context is the action-effect interval set by the experimental program. From this we can compute the degree to which the temporal distance between action and effect was perceived as shortened in the operant condition as compared to the baseline condition, i.e., the individual coupling strength ($\lambda_{individual}$), which was computed participant wise and for each delay condition

$$\lambda_{\mathrm{individual}} = \mathrm{Temporal\,Binding}/\mathrm{Baseline\,Action\,Effect\,Delay} \tag{7}$$

We can now aggregate the mean coupling strength (λ_{mean_pp}) over participants to extract an approximation of the perception of the situation. This is computed for all delay conditions separately

$$\lambda_{\mathrm{mean\_pp}} = Mean(\lambda_{\mathrm{individual}}) \tag{8}$$

In this model, the coupling strength (λ_{mean_pp}) should equate to the combined weights of action and effect. This is because, in this context, they are intended to collectively represent the total attraction between action and effect.

$$\lambda_{\mathrm{mean\_pp}} = w_{\mathrm{action}} + w_{\mathrm{effect}} \tag{9}$$

The SOCI coupling model suggests that even though the mutual attraction between cues remains similar to the one suggested in the SOCI fusion model, compared to it there is an overall increase in uncertainty about the action-effect relationship. Consequently, the model makes adjustments to the weights using the parameter $\sigma^2{}_{prior}$ as follows

$$w_{\mathrm{action}} = \frac{\sigma^2{}_{\mathrm{bl\_effect}}}{\sigma^2{}_{\mathrm{bl\_action}} + \sigma^2{}_{\mathrm{bl\_effect}} + \sigma^2{}_{\mathrm{prior}}} \tag{10}$$

$$w_{\mathrm{effect}} = \frac{\sigma^2{}_{\mathrm{bl\_action}}}{\sigma^2{}_{\mathrm{bl\_action}} + \sigma^2{}_{\mathrm{bl\_effect}} + \sigma^2{}_{\mathrm{prior}}} \tag{11}$$

The parameter $\sigma^2{}_{prior}$ represents the uncertainty about the action effect relation. This parameter is derived from the observed coupling strength and can be estimated from (9), (10) and (11), given

$$\lambda_{\mathrm{mean\_pp}} = \frac{\sigma^2{}_{\mathrm{bl\_effect}} + \sigma^2{}_{\mathrm{bl\_action}}}{\sigma^2{}_{\mathrm{bl\_action}} + \sigma^2{}_{\mathrm{bl\_effect}} + \sigma^2{}_{\mathrm{prior}}} \tag{12}$$

$$\sigma^2{}_{\mathrm{prior}} = \frac{(\sigma^2{}_{\mathrm{bl\_action}} + \sigma^2{}_{\mathrm{bl\_effect}}) - \lambda_{\mathrm{mean\_pp}}(\sigma^2{}_{\mathrm{bl\_action}} + \sigma^2{}_{\mathrm{bl\_effect}})}{\lambda_{\mathrm{mean\_pp}}} \tag{13}$$

$$\sigma^2{}_{\text{prior}} = \frac{1-\lambda_{\text{mean\_pp}}}{\lambda_{\text{mean\_pp}}} \times \left(\sigma^2{}_{\text{bl\_action}} + \sigma^2{}_{\text{bl\_effect}}\right) \tag{14}$$

To calculate the statistically optimal estimates, baseline estimate of the action (μ_{bl_action}) and the effect $\left(\mu_{bl\_effect}\right)$ must again be aligned in one timeframe, dependent on the action effect delay in the operant condition (see above). Using the weights as calculated above, the statistically optimal estimates (μ_{SO}) following the coupling model for action and effect are

$$\mu_{\text{SO\_action}} = \left((1 - w_{\text{effect}}) \times \mu_{\text{bl\_action}}\right) + \left(w_{\text{effect}} \times \mu_{\text{bl\_effect}}\right) \tag{15}$$

$$\mu_{\text{SO\_effect}} = \left(w_{\text{action}} \times \mu_{\text{bl\_action}}\right) + \left((1 - w_{\text{action}}) \times \mu_{\text{bl\_effect}}\right) \tag{16}$$

Note, that μ_{bl_effect} and λ_{mean_pp} are adjusted for each delay condition.

In cases of incomplete binding, the uncertainty surrounding the action-effect relationship limits the advantage that can be gained through multisensory integration. In other words, the minimal variance attainable through optimal maximum-likelihood estimation is modified by the coupling prior.

$$\sigma^2{}_{\text{min\_action}} = \frac{\sigma^2{}_{\text{bl\_action}} \times \left(\sigma^2{}_{\text{bl\_effect}} + \sigma^2{}_{\text{prior}}\right)}{\sigma^2{}_{\text{bl\_action}} + \sigma^2{}_{\text{bl\_effect}} + \sigma^2{}_{\text{prior}}} \tag{17}$$

$$\sigma^2{}_{\text{min\_effect}} = \frac{\sigma^2{}_{\text{bl\_effect}} \times \left(\sigma^2{}_{\text{bl\_action}} + \sigma^2{}_{\text{prior}}\right)}{\sigma^2{}_{\text{bl\_action}} + \sigma^2{}_{\text{bl\_effect}} + \sigma^2{}_{\text{prior}}} \tag{18}$$

These calculations for μ_{SO} and $\sigma^2{}_{\text{min}}$ are equal to those from the fusion model, when coupling strength (λ) approaches 1 or respectively the uncertainty of the prior ($\sigma^2{}_{\text{prior}}$) approaches 0.

Appendix B. Pilot experiment

In a short pilot experiment, N = 7 participants were asked to perform keypresses and thereby elicit a slightly delayed tone. The delay was specified by the program between 0 and 220 ms in 30 ms increments. We also measured the actual delay, which was subject to refresh rate etc. on participants end device. It was on average 6 ms longer than the specified delay. We entered the measured delay into the analysis. After each action-effect episode, participants were asked to rate whether they perceived the keypress and the tone to have been simultaneous. In the experiment, we piloted two different tone lengths, 50 ms (short) and 200 ms (long). We performed a logistic regression for each participant and finally computed a mean resultant regression function from the mean parameters of the participants. The 50 ms tone length led to a slightly smaller simultaneity threshold, which is why we chose to use the short tone in the main experiment.

$$\text{Relative probability of perceived simultaneity} = \frac{1}{1 + e^{3.08 + (-0.02 \times \text{delay})}}$$

The resultant regression model crosses the 50 % threshold at 150.94 ms measured delay. We therefore chose a 150 ms action-effect delay condition to maximize interindividual differences regarding perceived simultaneity.

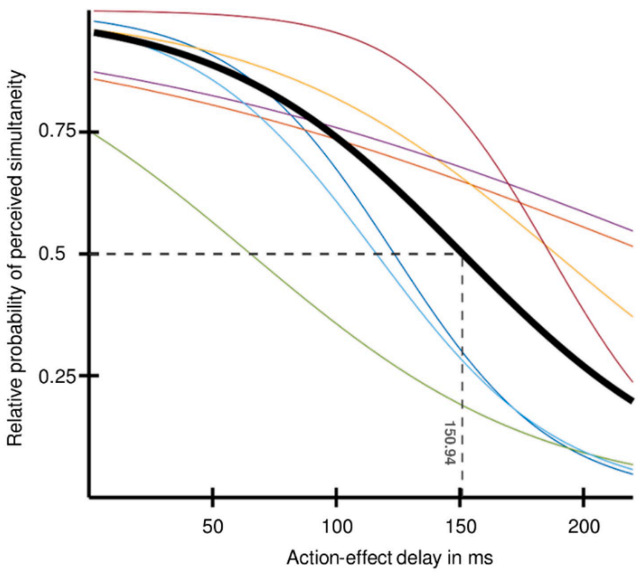

Fig. B1. Individual and mean regression functions
Note. Thin, colored lines indicate regression functions of single participants whereas the bold line indicates the regression function derived by averaging all participants function parameters. The action effect delay is given in measured delay, which was on average 6 ms larger than the delay specified by the program.

Appendix C. Self-reported nationality of all participants.

Experiment 1			**Experiment 2**		
Country	nationality	residence	Country	nationality	residence
Albania	1		France	1	
Estonia	1	1	Greece	1	1
Finland	1	1	Hungary	1	1
Greece	3	4	Italy	3	3
Hungary	2	3	Mexico	3	3
Ireland		1	Poland	1	1
Israel	1	1	Portugal	3	3
Italy	6	6	South Africa	10	11
Jordan	1		Swaziland	1	
Mexico	1	1	Spain		1
Nigeria	1		UK	4	4
Pakistan	1		USA	1	1
Poland	7	7			
Portugal	4	4			
Scotland		1			
Slovenia	2	2			

(*continued on next page*)

(*continued*)

Experiment 1				Experiment 2		
Country		nationality	residence	Country	nationality	residence
	South Africa	1	1			
	Spain	2	2			
	UK	11	12			
	USA	2	1			

Appendix D. Boxplot graph for action and effect binding data

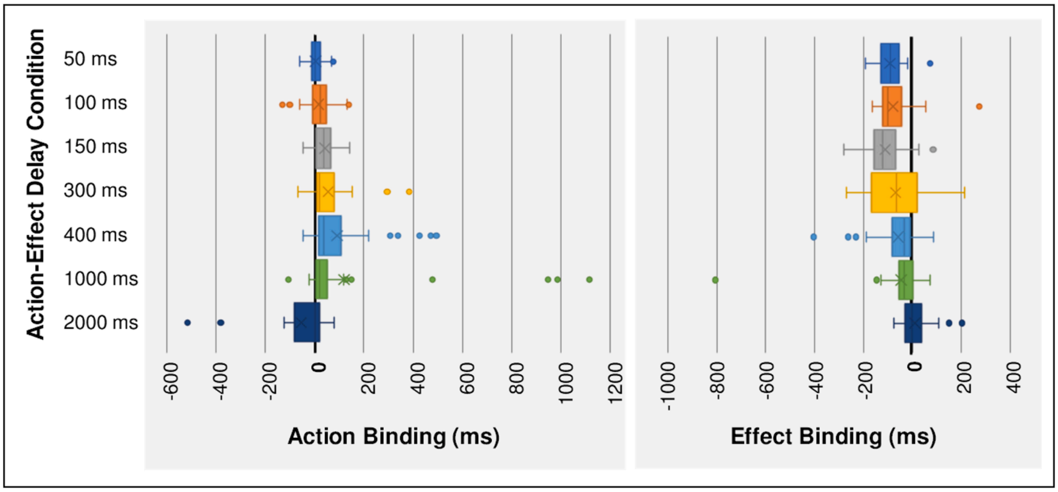

Fig. D1. Boxplot graph for action and effect binding data

Appendix E. Computation of −2 Log Likelihood values

First, we computed how likely each observed estimation error was according to SOCI models and comparison models in Likelihood values (LL):

$$LL = \frac{1}{\sigma_{\min} * \sqrt{2\pi}} \times \exp\left(-\frac{1}{2} \times \left(\frac{x - \mu_{SO}}{\sigma_{\min}}\right)^2\right)$$

x = Observed estimation error in the data

μ_{SO} = Predicted estimation error from the model.

σ_{min} = Predicted standard deviation

Likelihoods for the coupling model were assessed for actions and effects with μ_{SO} and σ_{min} adjusted to μ_{SO_action} or μ_{SO_effect} and σ_{min_action} or σ_{min_effect} respectively.

We then tr ansformed them into aggregate −2 Log Likelihood parameters individually for action and effect estimates and separately for each participant delay condition, and model:

$$\sum -2 \times LN(LL)$$

See Fig. E1 for an overview over Log Likelihood value dispersion on the participant aggregate level.

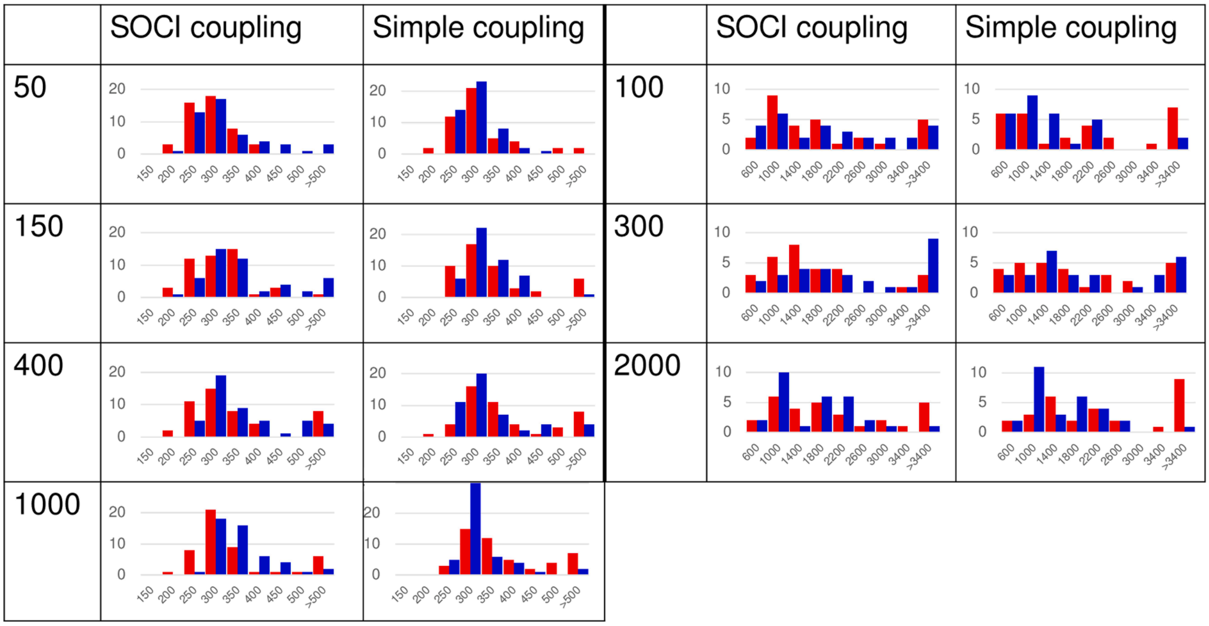

Fig. E1. Histogram data of likelihoods(−2LL) aggregated over trials.
Note. The categories display the upper bin boundary as the −2 Log Likelihood value, and the bar height represents the number of participants whose aggregate likelihood for this condition fell within this bin. Histograms for action estimates are depicted in red, while those for effect estimates are depicted in blue.

Appendix F. Correlational measures

F1. Methods

F1.1. Correlation of standardized binding and predicted weights

We computed the standardized action and effect binding in different conditions. That is, how much of the baseline action-effect interval (baseline action estimate to baseline effect estimate) was covered by action- and effect binding in the operant condition. In case of perfect model fit, the standardized action binding should correspond to the predicted weight of the effect and vice versa the standardized effect binding to the predicted weight of the action. Higher weights of a cue are related to stronger binding of all other cues towards it.

F1.2. Correlation of standardized measured and predicted variances

As a less conservative measure than exact similarity between minimal possible variances and observed variances we implemented the correlation of measured and predicted variances, each standardized by dividing it by the baseline variance.

F1.3. Intercorrelation of standardized action and effect binding

To test whether action binding and effect binding were interrelated independently from actual binding strength, we correlated the standardized action and effect binding scores with each other (standardized by the baseline action-effect interval, see above, 2.4.3). Multisensory integration would predict a negative correlation of the two, as increased dominance of one signal should always lead to decreased dominance of all others.

F2. Results

F2.2. Correlation of standardized binding and predicted weights

Predicted weights and standardized binding correlated significantly in the 100 ms condition. Both, the effect weight correlated with the standardized action binding, $r(27) = .52$, $p = .004$, and the action weight correlated with the standardized effect binding $r(27) = .44$, $p = .018$. In the 400 ms condition, standardized action binding was related to predicted weights, $r(46) = .29$, $p = .045$. Additionally, in the Experiment 2 2000 ms condition, both estimation errors were correlated with predicted weights, standardized action binding, $r(27) = -.53$, $p = .003$, standardized effect binding, $r(27) = .39$, $p = .038$. Due to the finding of action-effect repulsion (rather than binding) in this condition, the correlations entail diverging interpretations. None of the other correlations reached significance, all $ps > .122$ (see Fig. F1).

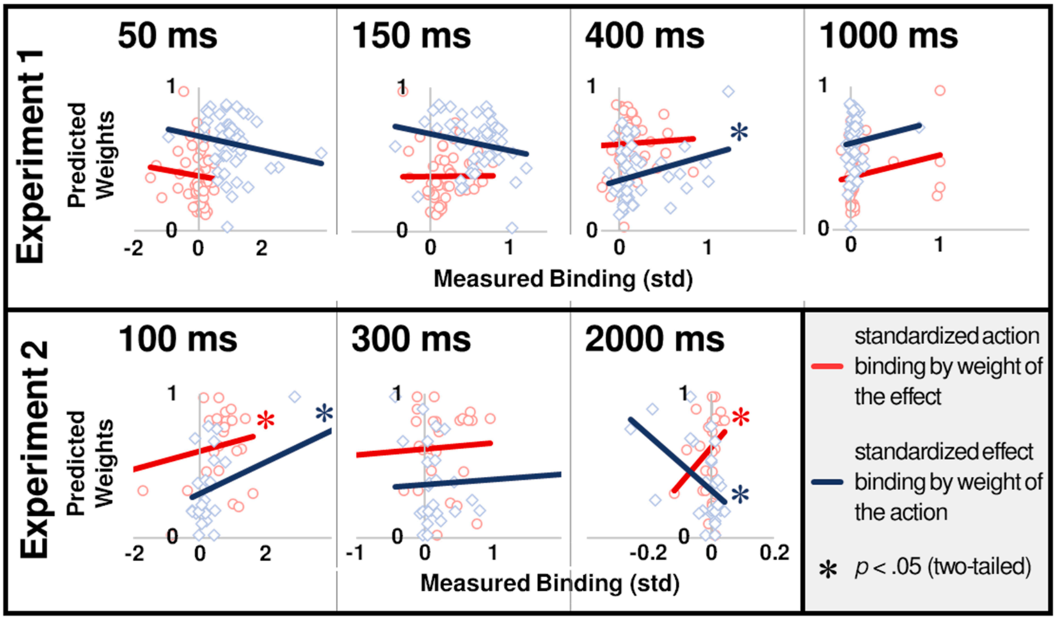

Fig. F1. Correlation of standardized binding and predicted weights
Note. Measured action (red) and effect binding (blue) was standardized by the baseline action-effect delay. They are each displayed against the weight they should correlate with positively.* p < .05.

F2.2. Correlation of standardized measured and predicted variances

Standardized observed variances of action estimates correlated with the predictions of the SOCI coupling model in the 50 ms condition, $r(46) = .59$, $p < .001$, 100 ms condition, $r(27) = .48$, $p = .008$, 150 ms condition, $r(46) = .35$, $p = .014$, and 300 ms condition, $r(27) = .69$, $p < .001$. The correlation of predicted and observed action variances in the 2000 ms condition was negative, $r(27) = -.47$, $p = .011$. Effect estimate variances were positively correlated with those predicted by the SOCI coupling model in the 50 ms condition, $r(46) = .50$, $p < .001$, 100 ms condition, $r(27) = .57$, $p = .001$, 150 ms condition, $r(46) = .46$, $p = .001$, and 400 ms condition, $r(46) = .38$, $p = .007$ (see Fig. F2).

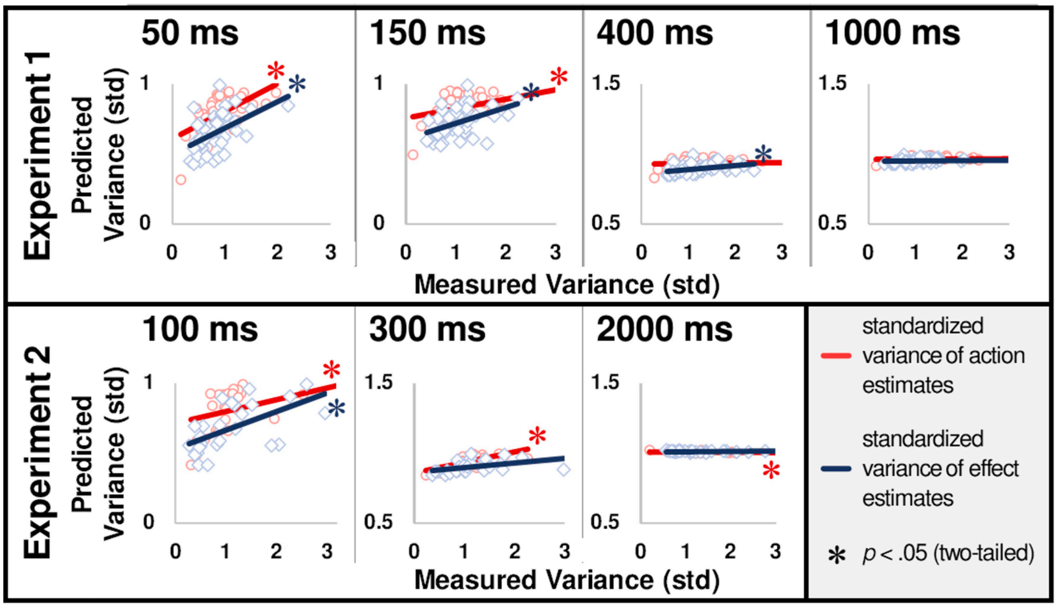

Fig. F2. Correlation of observed and predicted standardized variances of action and effect estimates
Note. Variance of operant action estimates by variance of predicted action estimates (red), both standardized by the variance of baseline action estimates; Variance of operant effect estimates by variance of predicted effect estimates (blue), both standardized by the variance of baseline effect estimates.

F2.3. Intercorrelation of standardized action and effect binding

When standardizing action binding and effect binding in relation to the overall binding score, they correlated negatively, as would be predicted by the rule of inverse effectiveness for the two shortest intervals, 50 ms: $r(46) = -.55$, $p < .001$; 100 ms: $r(27) = -.75$, $p < .001$, as well as for the 300 ms intervall $r(27) = -.82$, $p < .001$. This relationship did not reach significance in any of the other intervals, $ps > .160$ (see Fig. F3).

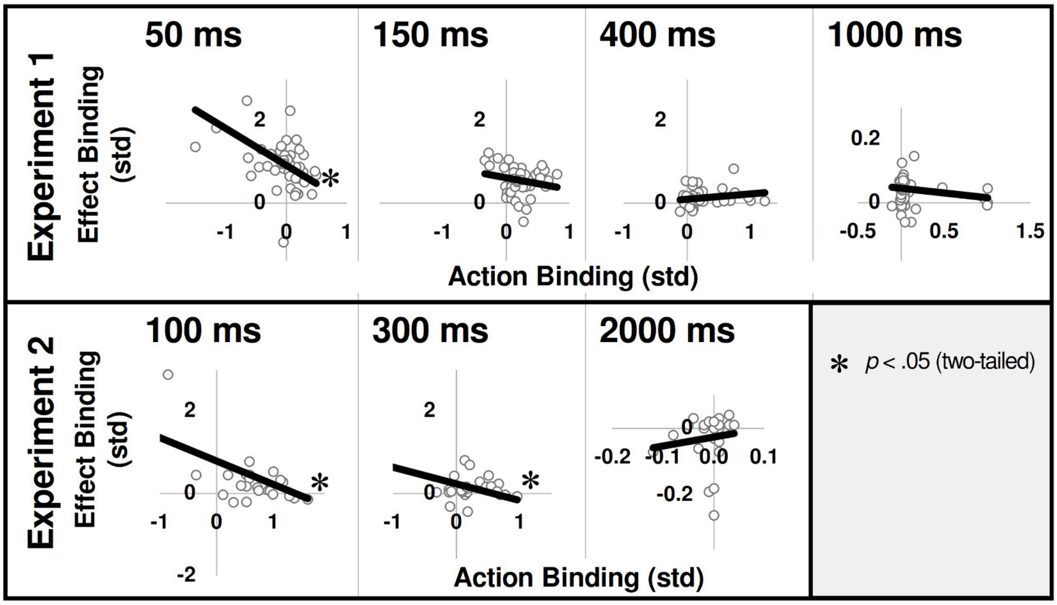

Fig. F3. Intercorrelation of standardized action and effect binding
Note. Measured action and effect binding were standardized by the baseline action-effect delay.
* $p < .05$.

Appendix G. Agnostic neural networks

To analyze the influence of the different factors independent from any specific model, we trained an unbiased black-box Neural Network (IBM SPSS Statistics 28, Multilevel Perceptron) with data from four published studies on temporal binding that employed the Libet Clock procedure, featured baselines, action and effect binding and where data was available to us (Kirsch et al., 2019; Klaffehn et al., 2021; Ruess et al., 2017; Schwarz et al., 2019). We trained the neural network on the data from these studies to predict the current data presented in this paper. A random half of the current data was used as “testing” sample used to track errors during training and prevent overtraining. This increases network training efficiency (IBM SPSS Statistics, 2021). The remaining half was used as “holdout” sample i.e., data the final model was tested against. Relative error for the holdout sample is therefore particularly indicative of general model performance. On basis of 50 training and testing sessions, we determined the average importance of each of the entered predictors for each of the operant parameters. As predictors, we entered the factors relevant for Multisensory Integration and the same factors that also play a role in the SOCI model, namely baseline estimates of actions and effects, variances of both these estimates and, finally, the true delay between action and effect. During the model generation, these predictors were only entered as standardized z-values.

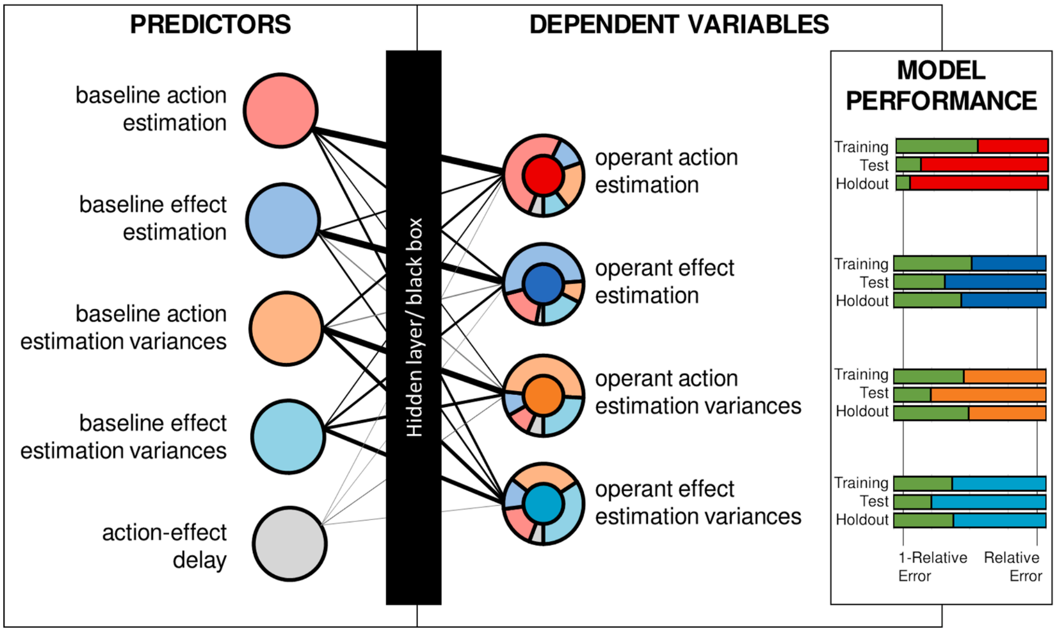

Fig. G1. Results of the black box neural network predicting temporal binding parameters
Note. Graphical representation of the model results. On the left, the predictive power of the predictors for the dependent variables is represented by the thickness of the lines and the colors in the outer rings around the dependent variable blobs. Model performance shows the relative error in the training, testing and holdout samples after training was complete.

Not surprisingly, the most important factors influencing the estimates were the respective similar parameters in the baseline conditions (See Fig. G1). However, while operant estimates were dominated by the corresponding baseline estimate and its variance, operant variances were predicted by both baseline variances and only to small degrees by estimates. The action-effect delay played a surprisingly negligible role in the neural network model, by making up only 3–7 % of the predictions. The largest influence was onto operant action estimate variances, which, albeit anecdotal, matches the earlier made point that delayed action estimates due to longer delays may taint data. Note here, that the data of the current experiments were not entered into the training sample, that generated the impact values. Therefore, this can be treated as a separate piece of evidence to this fact. Additionally, due to this division, the small impact of delay has not been driven by the uncommonly large action-effect intervals in the current experiments. This suggests that delay is not as powerful a predictor for temporal binding as commonly assumed.

Although no suitable model was at hand to be analyzed for comparison, the explanatory power of the neural network model does not appear to be exceptionally good. However, since the training sample was limited and it was tested on yet another sample, that included unlikely delay conditions, the absolute strength of the predictions is of little consequence. Yet, when predictive performance is compared between all three operant conditions the model performance for action estimates clearly stands out in that these predictions for the current dataset were worst. It is, therefore, reasonable to assume that action estimates if not also the other operant parameters, are notably influenced by another factor that was not represented as predictor in the model.

References

Bedard, G., & Barnett-Cowan, M. (2016). Impaired timing of audiovisual events in the elderly. *Experimental Brain Research, 234*(1), 331–340. https://doi.org/10.1007/s00221-015-4466-7

Buehner, M. J. (2012). 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

Chen, Y.-C., Lewis, T. L., Shore, D. I., & Maurer, D. (2017). Early Binocular Input Is Critical for Development of Audiovisual but Not Visuotactile Simultaneity Perception. *Current Biology, 27*(4), 583–589. https://doi.org/10.1016/j.cub.2017.01.009

Cramér, H. (1946). A contribution to the theory of statistical estimation. *Scandinavian Actuarial Journal, 1946*(1), 85–94. https://doi.org/10.1080/03461238.1946.10419631

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods, 47*(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y

Debats, N. B., Ernst, M. O., & Heuer, H. (2017). Perceptual attraction in tool use: Evidence for a reliability-based weighting mechanism. *Journal of Neurophysiology, 117*(4), 1569–1580. https://doi.org/10.1152/jn.00724.2016

Desantis, A., Roussel, C., & Waszak, F. (2011). On the influence of causal beliefs on the feeling of agency. *Consciousness and Cognition, 20*(4), 1211–1220. https://doi.org/10.1016/j.concog.2011.02.012

Ernst, M. O. (2006). A Bayesian view on multimodal cue integration. In G. Knoblich, I. M. Thornton, M. Grosjean, & M. Shiffrar (Eds.), *Advances in visual cognition. Human body perception from the inside out* (pp. 105–131). Oxford Univ. Press.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415*(6870), 429–433. https://doi.org/10.1038/415429a

Fereday, R., Buehner, M. J., & Rushton, S. K. (2019). The role of time perception in temporal binding: Impaired temporal resolution in causal sequences. *Cognition, 193*, Article 104005. https://doi.org/10.1016/j.cognition.2019.06.017

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

Holmes, N. P. (2009). The principle of inverse effectiveness in multisensory integration: Some statistical considerations. *Brain Topography, 21*(3–4), 168–176. https://doi.org/10.1007/s10548-009-0097-2

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

IBM SPSS Statistics (Version 28.0.1.0 (142)) [Computer software]. (2021). IBM Corp.

Ivanof, B. E., Terhune, D. B., Coyle, D., Gottero, M., & Moore, J. W. (2021). Examining the effect of Libet clock stimulus parameters on temporal binding. *Psychological Research, 86*(3), 937–951. https://doi.org/10.1007/s00426-021-01546-x

Kirsch, W., Kunde, W., & Herbort, O. (2019). Intentional binding is unrelated to action intention. *Journal of Experimental Psychology, 45*(3), 378–385. https://doi.org/10.1037/xhp0000612

Klaffehn, A. L., 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, and Psychophysics, 83*(8), 3135–3145. https://doi.org/10.3758/s13414-021-02314-0

Kuroki, D. (2021). A new jsPsych plugin for psychophysics, providing accurate display duration and stimulus onset asynchrony. *Behavior Research Methods, 53*(1), 301–310. https://doi.org/10.3758/s13428-020-01445-w

Langer, E. J. (1975). The illusion of control. *Journal of Personality and Social Psychology, 32*(2), 311–328. https://doi.org/10.1037/0022-3514.32.2.311

Legaspi, R., & Toyoizumi, T. (2019). A Bayesian psychophysics model of sense of agency. *Nature Communications, 10*(1), 4250. https://doi.org/10.1038/s41467-019-12170-0

Lewandowsky, S., & Farrell, S. (2011). *Computational modeling in cognition: Principles and practice*. Sage Publications.

Lush, P., Roseboom, W., Cleeremans, A., Scott, R. B., Seth, A. K., & Dienes, Z. (2019). Intentional binding as Bayesian cue combination: Testing predictions with trait individual differences. *Journal of Experimental Psychology. Human Perception and Performance, 45*(9), 1206–1217. https://doi.org/10.1037/xhp0000661

McGovern, D. P., Burns, S., Hirst, R. J., & Newell, F. N. (2022). Perceptual training narrows the temporal binding window of audiovisual integration in both younger and older adults. *Neuropsychologia, 173*, Article 108309. https://doi.org/10.1016/j.neuropsychologia.2022.108309

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*(5588), 746–748. https://doi.org/10.1038/264746a0

Noel, J.-P., de Niear, M., van der Burg, E., & Wallace, M. T. (2016). Audiovisual Simultaneity Judgment and Rapid Recalibration throughout the Lifespan. *PloS ONE, 11*(8), e0161698.

Parise, C. V., Spence, C., & Ernst, M. O. (2012). When correlation implies causation in multisensory integration. *Current Biology, 22*(1), 46–49. https://doi.org/10.1016/j.cub.2011.11.039

Ren, Y., Ren, Y., Yang, W., Tang, X., Wu, F., Wu, Q., Takahashi, S., Ejima, Y., & Wu, J. (2018). Comparison for younger and older adults: Stimulus temporal asynchrony modulates audiovisual integration. *International Journal of Psychophysiology, 124*, 1–11. https://doi.org/10.1016/j.ijpsycho.2017.12.004

Ruess, M., Thomaschke, R., & Kiesel, A. (2017). The time course of intentional binding. *Attention, Perception, and Psychophysics, 79*(4), 1123–1131. https://doi.org/10.3758/s13414-017-1292-y

Ruess, M., Thomaschke, R., & Kiesel, A. (2020). Intentional binding for unintended effects. *Timing & Time Perception, 1–9*. https://doi.org/10.1163/22134468-bja10005

Saad, L., Musolino, J., & Hemmer, P. (2022). Bayesian rational memory model simulates temporal binding effect. *Proceedings of the 44th Annual Meeting of the Cognitive Science Society*, *8*(44), 3549–3556.

Saad, L., Hemmer, P., & Musolino, J. (2022). Rethinking temporal binding. *Advance online publication*. https://doi.org/10.31234/osf.io/taejk

Sanders, P., Thompson, B., Corballis, P. M., & Searchfield, G. (2019). On the timing of signals in multisensory integration and crossmodal interactions: A scoping review. *Multisensory Research, 32*(6), 533–573. https://doi.org/10.1163/22134808-20191331

Schwarz, K. A., & Weller, L. (2023). Distracted to a fault - Attention, actions, and time perception. *Attention, Perception, and Psychophysics, 85*, 301–314.

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

Shams, L., Kamitani, Y., & Shimojo, S. (2000). What you see is what you hear. *Nature, 408*(6814), 788. https://doi.org/10.1038/35048669

Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: Current issues from the perspective of the single neuron. *Nature Reviews. Neuroscience, 9*(4), 255–266. https://doi.org/10.1038/nrn2331

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*, Article 103219. https://doi.org/10.1016/j.concog.2021.103219

Venskus, A., Ferri, F., Migliorati, D., Spadone, S., Costantini, M., & Hughes, G. (2021). Temporal binding window and sense of agency are related processes modifiable via occipital tACS. *PloS ONE, 16*(9), e0256987.

Wallace, M. T., Roberson, G. E., Hairston, W. D., Stein, B. E., Vaughan, J. W., & Schirillo, J. A. (2004). Unifying multisensory signals across time and space. *Experimental Brain Research, 158*(2), 252–258. https://doi.org/10.1007/s00221-004-1899-9

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

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*(3), 467–474. https://doi.org/10.1007/s00221-013-3419-2

Zachariah, D., & Stoica, P. (2015). Cramér-Rao bound analog of Bayes' rule [Lecture notes]. *IEEE Signal Processing Magazine, 32*(2), 164–168.