



OPEN

Evaluating the Bayesian causal inference model of intentional binding through computational modeling

Takumi Tanaka

Intentional binding refers to the subjective compression of the time interval between an action and its consequence. While intentional binding has been widely used as a proxy for the sense of agency, its underlying mechanism has been largely veiled. Bayesian causal inference (BCI) has gained attention as a potential explanation, but currently lacks sufficient empirical support. Thus, this study implemented various computational models to describe the possible mechanisms of intentional binding, fitted them to individual observed data, and quantitatively evaluated their performance. The BCI models successfully isolated the parameters that potentially contributed to intentional binding (i.e., causal belief and temporal prediction) and generally better explained an observer's time estimation than traditional models such as maximum likelihood estimation. The estimated parameter values suggested that the time compression resulted from an expectation that the actions would immediately cause sensory outcomes. Furthermore, I investigated the algorithm that realized this BCI and found probability-matching to be a plausible candidate; people might heuristically reconstruct event timing depending on causal uncertainty rather than optimally integrating causal and temporal posteriors. The evidence demonstrated the utility of computational modeling to investigate how humans infer the causal and temporal structures of events and individual differences in that process.

Time and causality are intertwined in the brain. Temporal relationships between events tell us about the causal structures in the world^{1,2}, whereas causal belief often distorts our subjective time^{3–7}. A well-known causality-driven illusion is intentional binding—the compression of the time interval between an action and its consequence. In a typical method⁸, this effect is assessed by comparing two conditions. In the operant condition, one's action (e.g., keypress) causes the presentation of a stimulus (e.g., tone). In the baseline condition, the same events occur by themselves in separate trials. Consequently, the observers perceive the action to be performed later and the stimulus to be presented earlier in the operant condition than in the baseline one. Named *intentional binding*, this effect is originally thought to be specific to (or at least prominent in) intentional action and its sensory outcome⁹. When the movement was triggered by an external magnetic stimulation to the brain, Haggard et al.⁸ observed the repulsion, not compression, of the perceived interval. Moreover, compression was no longer observed when the observer was led to believe that there was no causality between the two events¹⁰. Intentional binding therefore has been utilized as a proxy for the sense of agency¹¹; however, its interpretation has sparked considerable debate. Recent studies have questioned the necessity of intentionality in this effect. Similar temporal compression has been observed for outcomes of externally induced^{12–14} or forced^{15,16} actions and for unintended outcomes¹⁷. Moreover, the essentiality of movement in intentional binding has been challenged by findings showing temporal attraction between externally triggered causal events^{5,18,19}, between the inhibition of an action and its consequences²⁰, and even between the actions of others and their outcomes²¹. These discoveries have prompted some scholars to propose that temporal compression may be indicative of a broader mechanism of causality perception, thus favoring the term “temporal binding” or “causal binding”^{5,18}. The connection between intentional binding and the sense of agency is still disputed, with some studies confirming their correlation²², while others do not²³. This ongoing debate highlights the need for elucidation of the fundamental nature and underlying mechanisms of intentional binding^{24,25}.

Multisensory integration framework has attracted growing attention as a potential explanation for intentional binding^{26,27}. Humans are consistently inundated with noisy signals from the external world. To optimize

Graduate School of Humanities and Sociology and Faculty of Letters, The University of Tokyo, Tokyo, Japan. email: kino31513@l.u-tokyo.ac.jp

perception in such a situation, the Bayesian brain integrates multiple signals weighted by their reliability. Computationally, this integration serves to minimize the variance of final estimates, known as Maximum-Likelihood Estimation (MLE)²⁸. In this context, intentional binding occurs when one estimates the timing of an event (e.g., action) by considering the timing of another event (e.g., tone). Supporting this notion, amplifying the sensory noise around motor or outcome events, thereby reducing their reliability, resulted in the timing estimates of these events being more heavily biased towards the timing of the other, more reliable event^{26,29–31}. However, MLE explains only how, but not when, signals are integrated. Importantly, the perceptual system cannot unconditionally merge available signals because integrating irrelevant signals worsens estimation. Addressing this problem, the recent Bayesian Causal Inference (BCI) model elucidates the cognitive processes by which the brain infers the causal structure—determining which signals originate from the same source—through the synthesis of prior knowledge and sensory evidence^{32–34}. When one detects a remarkable (temporal or spatial) discrepancy between events, it breaks down the perceived causality and thereby integration³⁵.

Based on the BCI model, Legaspi and Toyozumi³⁶ implemented a computational model of intentional binding. In the multisensory contexts, integration is generally driven by the *unity prior* that different signals should represent an identical source in time and space^{37,38}. This assumption may not be appropriate in the case of intentional binding where the action and outcome are usually separated by a detectable (typically 250 ms) gap²⁹. Therefore, Legaspi and Toyozumi³⁶ generalized the model by introducing a free parameter, μ_{AO} , that represents the expectation for the interval length between the *Action* and *Outcome*. An ideal observer represents the probability of experiencing a certain action-outcome interval, which biases temporal estimation. Such a flexible prior for temporal correlations is called *the coupling prior*^{37,38}, as opposed to the unity prior for simultaneity. Importantly, their model predicts either repulsion or compression depending on whether the interval between sensory signals is shorter or longer than the prior. Thus, their model can explain the repulsion in an involuntary situation as well as compression⁸. I refer to this model as the Legaspi and Toyozumi (LT) model in the current study.

Computational modeling has multiple benefits³⁹. It outputs probabilistic predictions of an observer's responses (e.g., time estimates) given sensory inputs. Comparing these predictions with observed data allows one to quantitatively evaluate the model's ability. One can also compare the performance of different models and gain insight into the internal processes that produce the responses. Furthermore, estimating the model parameters specifies (the source of) interindividual and intergroup variation. While intentional binding is generally a robust phenomenon, large individual differences are linked to specific psychiatric disorders such as schizophrenia^{40–42}. Examining internal parameters may provide cues on what characterizes one's binding tendency.

Despite these potential advantages, the LT model's validity has not been sufficiently examined thus far. Legaspi and Toyozumi³⁶ reproduced tendencies reported in two significant studies^{8,31} but did not fit their model to observations at either individual or trial-by-trial levels. Moreover, although Legaspi and Toyozumi³⁶ validated their model by referring to the involuntary condition in⁸, it is unclear what process the external stimulation modulated, e.g., attention, motor execution, or causal cognition. It is thus worth testing the LT model in a typical voluntary condition. Finally, they considered only one model (i.e., the LT model) and did not compare it with other types of BCI models with different algorithms or those with non-Bayesian alternatives.

Therefore, the current study quantitatively evaluated the computational models of intentional binding based on empirical data. I measured intentional binding using a classic clock method. Instead of employing an atypical involuntary condition, I manipulated signal intervals by varying the physical size of action-outcome gaps. As mentioned above, the LT model that is based on a coupling prior predicts repulsion rather than compression when the signal interval is sufficiently short. Technically, this manipulation was required to isolate the causal and temporal priors, which both influence the binding effects³⁶. Implementing various computational models, I first examined their ability to distinguish their free parameters via parameter recovery. Then, for the first time, I fitted the models to observed data. Based on the model fit and estimated parameter values, I discuss the possible mechanisms underlying intentional binding.

Results

Behavioral results

I conducted an online experiment (N=76) based on a typical intentional binding task (Libet clock paradigm; see “Method” for details). The experiment was developed based on the JavaScript plugin with which Galang et al.⁴³ replicated robust intentional binding. In separate blocks, observers performed arbitrary keypress and/or heard a tone while viewing a clock rotating on a screen, and they estimated the timing of either of the events based on the clock-hand position. In the baseline tasks, they estimated the timing of the keypress or tone presentation that occurred independently. The operant tasks included three conditions of the action-outcome interval; observers' keypress randomly caused a tone after 0 ms, 250 ms, or 500 ms. Estimation errors were calculated by subtracting the actual time of a keypress or tone from the reported time of the corresponding event.

The distributions of the estimates are illustrated in Fig. 1a. The baseline estimates for the keypress and tone timing were respectively biased in anticipatory and lagged ways (keypress baseline: mean = -52.22, SD = 61.56; tone baseline: mean = 18.27 ms, SD = 61.56 ms). The estimation errors from the analysis of variance (ANOVA) revealed the main effects of the conditions (baseline, 0 ms operant, 250 ms operant, or 500 ms operant) for both keypress ($F(3, 225) = 16.04$, $MSE = 1029.85$, $p < 0.001$, $\eta_G^2 = 0.028$) and tone timing ($F(3, 225) = 10.66$, $MSE = 2, 855.29$, $p < .001$, $\eta_G^2 = 0.030$). Compared to the baseline condition, the keypress timing was overestimated in the operant trials with the 250 ms ($t(75) = 4.59$, $p < 0.001$, $d = 0.30$) and 500 ms ($t(75) = 3.94$, $p = 0.001$, $d = 0.31$) action-outcome intervals but not in the 0 ms ($t(75) = -0.21$, $p = 0.833$, $d = 0.01$) interval. There were also significant differences between 0 and 250 ms ($t(75) = 6.63$, $p < 0.001$) and 0 ms and 500 ms ($t(75) = 4.23$, $p < 0.001$), but not between 250 and 500 ms ($t(75) = 1.30$, $p = 0.394$). In contrast, the tone timing was underestimated compared to the baseline in all the interval conditions (0 ms: $t(75) = -4.54$, $p < 0.001$, $d = 0.38$; 250 ms: $t(75) = -3.60$,

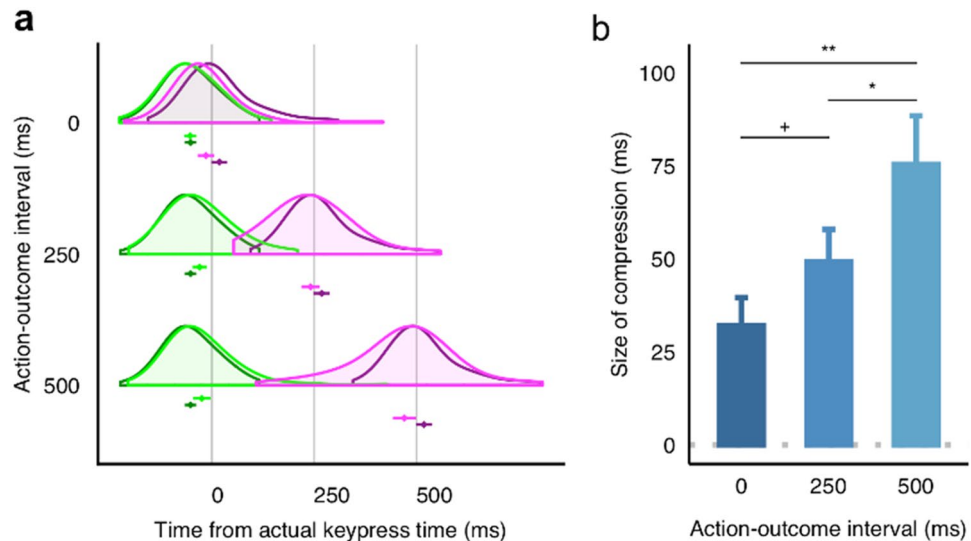


Figure 1. Behavioral results of the experiment. **(a)** The distributions and means of the time estimates (the dots with error bars below the distribution plots) were reported in each action-outcome interval condition (0, 250, or 500 ms). The x-axis represents the timeline, with the actual time of keypress marked as zero. The time estimates are represented by color-coded plots: green for keypress estimates and purple for tone estimates. Lighter shades of these colors represent estimates in the operant condition, while darker shades denote baseline estimates. For ease of comparison, the single baseline estimates for keypress and tone are replicated across each interval condition. This is done by adjusting their offsets, allowing for a direct visual comparison with the operant condition estimates in each interval. **(b)** The mean magnitude of the time compression (i.e., intentional binding) as a function of action-outcome interval length. The sizes of the compressions were calculated by summing up the perceptual shifts from the baseline for the keypress and for the tone (see Methods for details). The symbols above the pairs of bars indicate significance in the post-hoc pairwise *t*-tests: ** $p < 0.01$, * $p < 0.05$, and + $p < 0.10$. All error bars represent standard errors for within-participants design⁴⁴.

$p = 0.002$, $d = 0.29$; and 500 ms: $t(75) = -4.76$, $p < 0.001$, $d = 0.42$). While there was no significant difference in the timing estimations between the 0 ms and 250 ms conditions ($t(75) = 0.74$, $p = 0.464$), and between the 0 ms and 500 ms conditions ($t(75) = -1.25$, $p = 0.427$), a significant difference was observed between the 250 and 500 ms conditions ($t(75) = -3.65$, $p = 0.002$). In summary, I confirmed the biases in time estimation consistent with typical intentional binding (i.e., compression) in all the conditions except for the action in the 0 ms condition.

Summing up action and outcome binding, I plotted the overall intentional binding in Fig. 1b. Perceived intervals were compressed in proportion to the lengths of the physical intervals ($F(2, 150) = 7.53$, $MSE = 4,793.56$, $p = 0.001$, $\hat{\eta}_G^2 = 0.040$). Although the difference between the 0 and 250 ms conditions was not statistically significant ($t(75) = 1.77$, $p = 0.080$, $d = 0.24$), the 500 ms condition elicited greater compression than the 0 ms ($t(75) = 2.87$, $p = 0.011$, $d = 0.46$) and 250 ms ($t(75) = 3.43$, $p = 0.003$, $d = 0.23$) conditions. One-sample *t*-tests confirmed the occurrence of compression in all the intervals (0 ms: $t(75) = 4.27$, $p < 0.001$, $d = 0.48$; 250 ms: $t(75) = 5.60$, $p < 0.001$, $d = 0.64$; and 500 ms: $t(75) = 5.75$, $p < 0.001$, $d = 0.65$).

Computational modeling

Introduction of models

To explore the mechanism that produced these behavioral results, a computational modeling analysis was performed. Here, I explored twelve computational models as potential explanations for intentional binding. As the details for the model algorithms, fitting, and comparison are described in the “Methods” section, I briefly introduce the concepts of each model.

Eight of the twelve models were variations of the BCI accounts in which an observer infers the causal and temporal structures of events based on their prior expectations and sensory evidence. These models assume that observers have two priors regarding causal and temporal relationships between the action and the tone, respectively. The causal prior, $P(\xi = 1)$, represents the prior probability of a causal relationship between an action (i.e., keypress) and subsequent events (i.e., tone). The temporal prior depends on the causal belief. The observers have an expectation for the interval between causally linked keypress and tone, which plays a role as joint prior of keypress and tone timing. Here, this prior is assumed to be a Gaussian distribution with mean μ_{AO} and standard deviation (SD) σ_{AO} . For an observer with a unity prior, μ_{AO} approaches zero. Meanwhile, for the acausal events that occur independently, there should be no relationship between event timing and therefore the prior for event interval would be flat (i.e., a uniform distribution) instead of a Gaussian distribution. The likelihood is calculated as the probability at which the occurrence of an event is signaled to the brain at a given time. For causal and acausal cases, one can shape the posterior distribution for event timing by combining the temporal prior and the likelihood function and obtain the candidates of time estimation as the peak location of

the distribution (maximum-a-posteriori; MAP). Therefore, temporal estimation does not fully depend on when sensory signals arrive but is subject to their expected intervals for the likely causal scenario.

To draw final estimates, an observer can employ several strategies on how to utilize the two potential estimates depending on causality. The LT model³⁶ directly compares the posterior probabilities of estimates (i.e., MAP) between causal and acausal cases and identifies more likely estimates. This corresponds to adopting the MAP estimates of the joint posterior of causal scenario and event timing. In contrast, multisensory research has proposed that causal inference and timing estimation can be hierarchically performed; the brain first infers causality by calculating the marginal posterior probability and then explores the likely event timing accordingly⁴⁵. We consider three strategies respectively called *model-selection*, *model-averaging*, and *probability-matching*⁴⁶. Model-selection refers to a strategy where the observer selects the most likely causal scenario (either the presence or absence of a causal relationship) based on the posterior probabilities and then makes estimations in this chosen scenario. This is different from model-averaging, where the observer takes into account both causal and acausal scenarios, averaging the estimates across these scenarios weighted by their respective posterior probabilities. Probability-matching, on the other hand, is another strategy where the observer randomly chooses between the causal and acausal scenarios based on their posterior probabilities for each trial, thereby *matching* the probability distribution of the scenarios. In summary, the BCI variants comprised two components, i.e., two types of temporal prior (i.e., the unity and coupling priors) and four estimation strategies. Factorially combining them resulted in the eight BCI models.

In contrast to the BCI models, I expressed a mandatory integration model without causal inference by fixing $P(\xi = 1)$ to one. In this model, the signals are always integrated regardless of their timing, which means simple MLE. Because the four estimation strategies above make no difference when $P(\xi = 1) = 1$, there were only two types of mandatory integration models with the coupling and unity priors, respectively.

The remaining two models represent non-Bayesian accounts. First, I assessed a fixed-criterion model in which the observer provided temporal judgments by averaging the timing of the action and outcome whose temporal gap was smaller than a certain threshold ϕ ⁴⁷. Alternatively, the false-report model assumes that observers can falsely report the non-target event instead of the target event with a certain probability, P_f . For instance, when an observer estimates the outcome timing in some trials of the action operant condition, the mean action estimate apparently shifts toward the outcome, and vice versa. Although averaging the data including these erroneous responses can produce apparent binding-like results, such a risk has been overlooked in previous studies.

These 12 models above predict estimates for keypress and tone timing in operant tasks based on the distribution of the estimates in the baseline tasks. In the baseline tasks, as there is only one event present, the observers simply make time estimates based on the sensory inputs. Therefore, one can regard the distribution of time estimation in baseline tasks as the distribution of sensory noises. The mean and variance of this baseline distribution are included in all 12 models as the fixed parameters. Moreover, the models can be evaluated relative to a baseline null model in which observers derive estimates similarly in the operant and baseline tasks and thus no binding occurs.

Parameter recovery

First, I investigated whether the models could identify their parameters via parameter recovery. Each simulated model was fitted to the data for 76 ideal observers whose free parameters were randomly sampled from the possible ranges (see “Methods” for details). The original and recovered parameter values were generally robustly correlated (mean $r = 0.74$, $ts > 3.48$, $ps < 0.001$, see Fig. 2 for details). Except for the BCI models with model-selection, all the models recovered their parameters with high accuracy. The model-selection strategy always rejected the integration when the causal posterior was lower than a certain threshold, which seemed to make it difficult to isolate the parameters. For more detailed results, including all the values of original and recovered parameters, see Supplementary Table S1.

Model recovery

I also conducted a comprehensive model recovery analysis to assess the discriminative ability of the 12 computational models. The analysis involved fitting each model to datasets generated by these same models obtained in the earlier parameter recovery. The fitting was evaluated based on the Akaike information criterion (AIC), which considers both the model's goodness of fit and its complexity. Figure 3 presents a model recovery matrix, which visualizes the comparative effectiveness of each model in predicting the datasets generated by itself and by the other models. The heatmap revealed a general trend where models accurately identified data from their own simulations, as indicated by the prominent diagonal line. The analysis generally demonstrated good discrimination between Bayesian models and non-Bayesian models such as mandatory integration, fixed-criterion, and false-report. However, the matrix also revealed instances of less successful model discrimination. Especially, the Bayesian models with the joint posterior strategy and those with the model-selection strategy showed a less distinct separation, indicating a challenge in differentiating between the data they produced. The observed overlaps in discrimination underscore the need for careful interpretation in subsequent analyses.

Fitting results

The models were fitted to each observer's data at the individual level and to the pooled data at the group level. In each case, I estimated the best free parameters using MLE.

The best models for individual observers are shown in Fig. 4a. The AIC indicated that the BCI models best explained 56 out of 76 observers' data, whereas the fixed-criterion model yielded the best fits for only five observers (for all the AIC values for each model for each observer, see Supplementary Table S2). No observer was best explained by the mandatory integration models. At the individual level, there was no dominant BCI strategy

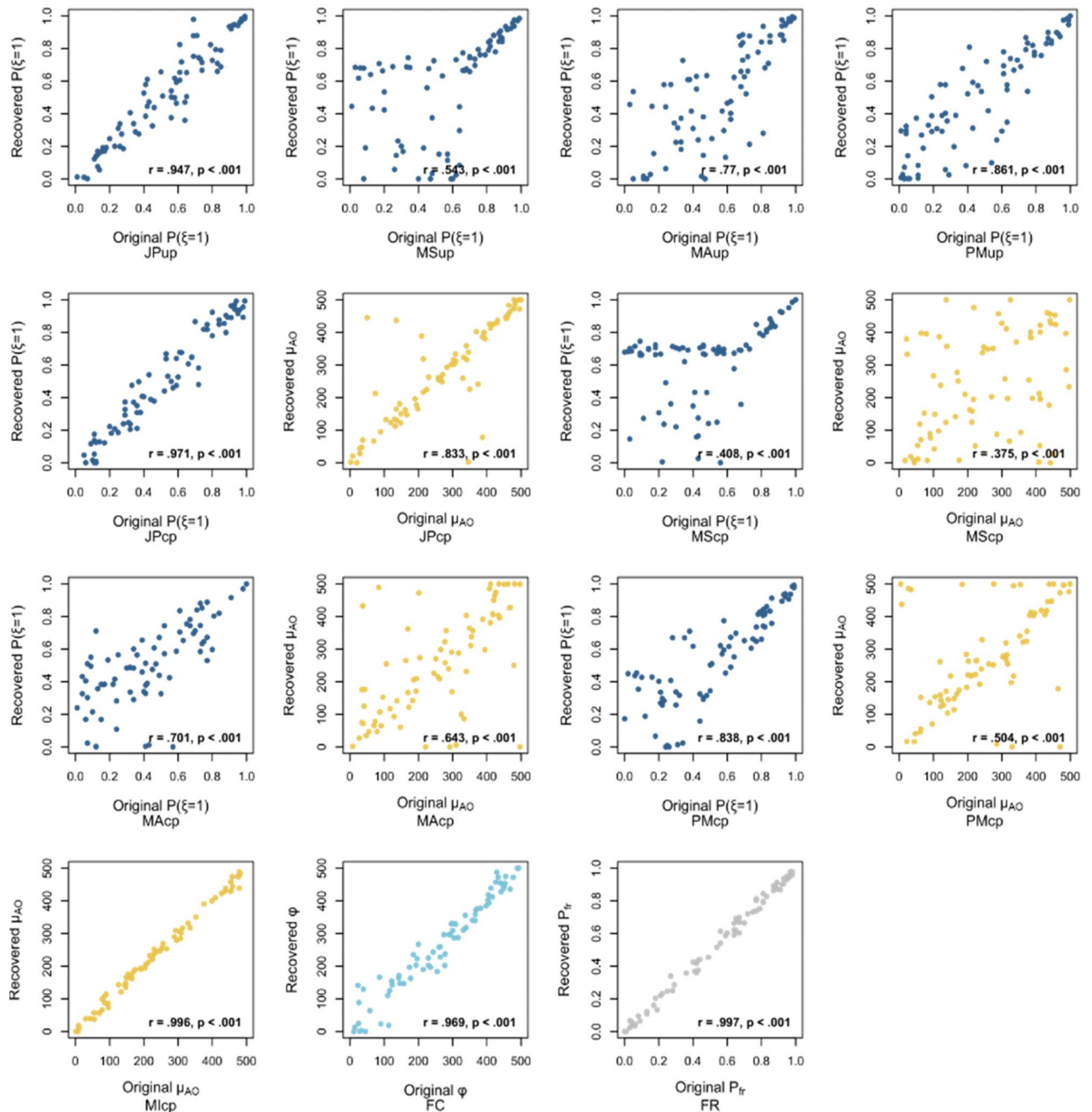


Figure 2. Results of parameter recovery. The relationships between original (randomly sampled) and recovered parameter values. Pearson's correlation coefficients, r , and p -values in the tests for no correlation are shown for each pair. The words under each plot indicate model names. The capital characters in the model names indicate the algorithms (JP: BCI with joint posterior; MS: BCI with model-selection; MA: BCI with model-averaging; PM: BCI with probability-matching; MI: mandatory integration; FC: fixed-criterion model; and FR: false-report model). For the BCI models, the lowercase characters indicate the type of temporal prior: up: unity prior; and cp: coupling prior.

shared by the observers. Of the 56 observers, 26 preferred the BCI models with the unity prior, and 30 preferred those with the coupling prior. Lumping different priors, 11 observers preferred the BCI models with the joint posterior, 22 the probability-matching, 19 the model-averaging, and 4 the model-selection. I plotted the model parameters' estimated values in Fig. 4b–e. The estimated $P(\xi = 1)$ tended to be smaller in the BCI models based on the joint posterior than in other BCI algorithms (Fig. 4b). While Legaspi and Toyozumi³⁶ assumed $P(\xi = 1)$ of 0.9 in the voluntary condition, the mean of the estimated value in the LT model (i.e., the BCI model based on the coupling prior and joint posterior) was ~ 0.5 . Moreover, the mean (\pm SD) of μ_{AO} was 95.66 (± 94.32) ms in the LT model and 117.44 (± 130.09) ms in the BCI models overall (Fig. 4c), which were also smaller than previously suggested³⁶. The mean (\pm SD) of estimated ϕ was 186.65 (± 104.03) in the fixed-criterion model (Fig. 4d).

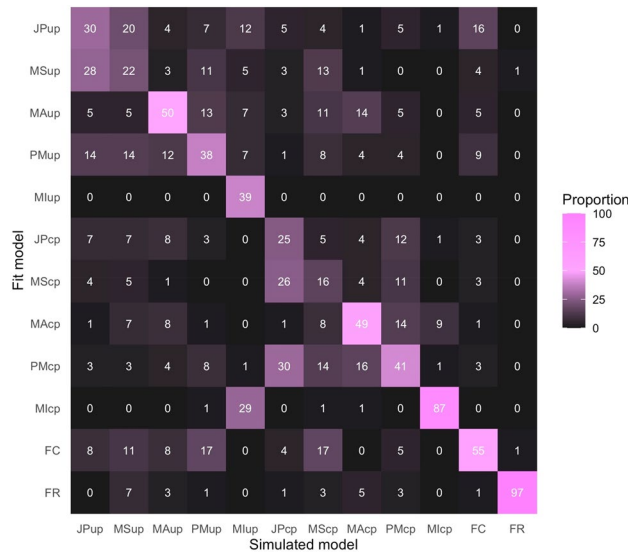


Figure 3. Model Recovery Matrix. This heatmap visualizes the cross-comparison of fit for 12 distinct models applied to datasets simulated by the same suite of models. Each cell indicates the percentage of datasets for which the row model provided the best explanation of the column-simulated data, with the color intensity reflecting the proportion (from 0 to 100, as denoted by the color scale on the right). Shades closer to pink signify a higher percentage, emphasizing the instances of accurate model recovery. The diagonal dominance suggests that models tend to best fit the data they generated, highlighting the discriminative power and fidelity of the modeling approach. See the Fig. 2 caption for the meaning of the model names’ acronyms.

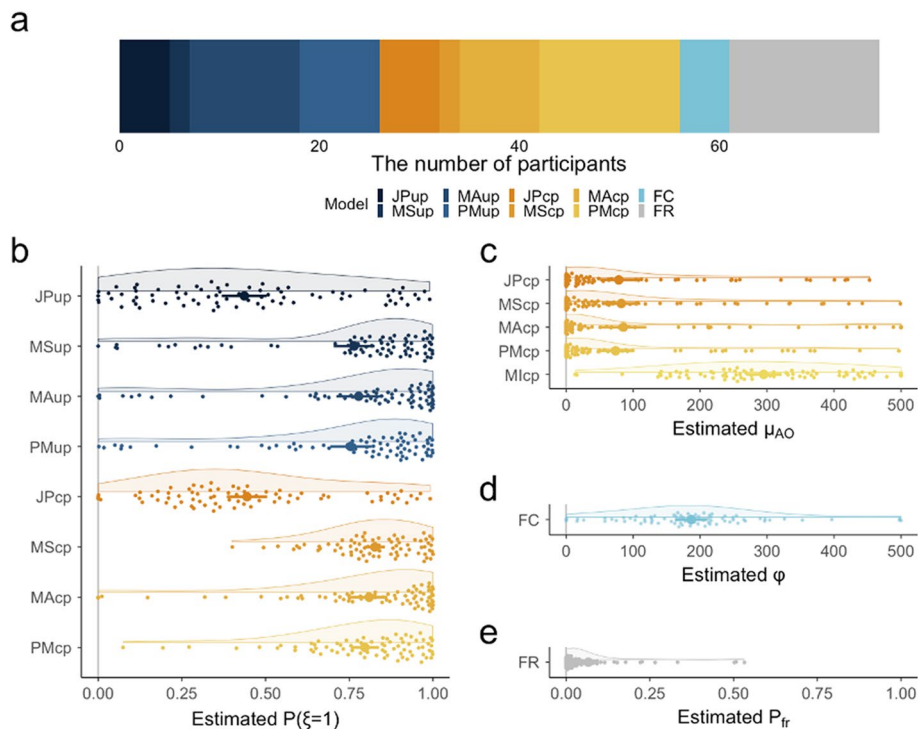


Figure 4. Results of individual-level model fitting. (a) The portfolio of models that best accounted for individual observers’ data (76 observers in total). (b) Distribution of estimated values of $P(\xi = 1)$ for each type of BCI model. Small dots correspond to individual observers and large dots correspond to mean values. Error bars represent standard errors for within-participants design⁴⁴. (c) Distribution of the estimated values of μ_{AO} for models with the coupling prior. (d) Distribution of the estimated values of ϕ in the fixed-criterion model. (e) Distribution of the estimated values of P_{fr} in the false-report model. See the Fig. 2 caption for the meaning of the model names’ acronyms.

Surprisingly, the remaining 15 observers' estimations were best accounted for by the false-report model (Fig. 4a). Probably due to these observers, the group-level analysis for all pooled data identified the false-report model as the best (see Supplementary Table S3 for detailed results). The model's best fitting parameter was $P_{fr} = 0.04$, implying that observers made erroneous estimates in ~4% of all the trials, on average (see also Fig. 4e).

Considering that the 15 observers whose estimates were best accounted for by the false-report model might not appropriately perform the task, as planned (see pre-registration), I conducted another group-level fitting to the data excluding them. The results of this fitting are shown in Table 1, including the estimated parameter values, log-likelihood (LL), and AIC for each candidate model. They reveal that the probability-matching BCI model with the unity prior and that with the coupling prior almost equally better account for the remaining data. According to Hilbe⁴⁸, an AIC difference of less than 2.5 can be interpreted as nonsignificant. With an AIC difference of ~10 from these two models, the BCI based on the joint posterior and fixed-criterion model yield the next-best fits. Notably, only the mandatory integration models showed lower performance than the null model.

Discussion

BCI has been considered a promising account for intentional binding. While the theory has recently been formulated as a computational model, it has not been thoroughly tested. In this study, I implemented a range of computational models, including both the BCI and non-BCI models, and quantitatively evaluated their ability. Instead of employing an involuntary condition, I manipulated the physical action-outcome interval length to test the model prediction. Although some partial problems remain, which will be discussed below, parameter recovery and model recovery ensured the validity of the model at a certain level. This allowed me to fit the models to observed data for the first time.

The mechanism underlying intentional binding

The parameter estimation and model comparison provided several implications for the underlying mechanism for intentional binding. Most of the observers were well explained by variants of BCI models, whereas some were explained by potential irregular responses. Notably, the BCI models outperformed the simple MLE as well as the fixed-criterion models at both individual and group levels. This indicates that humans do not always integrate an action and subsequent event but decide whether and how to integrate them by inferring their causal relationship.

I considered various BCI algorithms. While there was no consensus among individuals, probability-matching showed the best performance at the group level. In probability-matching, one sequentially infers the causal scenario and event timing, which is different from inference based on a joint posterior in the LT model³⁶. Interestingly, probability-matching is a computationally suboptimal strategy compared to model-averaging⁴⁶. As the calculation of the joint posterior and model-averaging are expensive, humans may heuristically reconstruct event timing depending on causal uncertainty. This is likely, given that probability-matching is known to be predominantly used in several perceptual and cognitive tasks^{49–53}. It should be noted, however, that some Bayesian model algorithms, such as the probability-matching model with the coupling prior and the LT model, did not discriminate each other very well (see Fig. 3). Refinement of the models and experiments with larger sample sizes and more repetitions may contribute to drawing more robust conclusions.

In addition to the causal prior, the BCI models have the temporal prior regarding the action-outcome interval. The estimated mean values of temporal prior (μ_{AO}) were distributed around zero, with some variance across individual observers (Fig. 4c). The group-level fitting also estimated it at close to zero (see Table 1); therefore, introducing the coupling prior did not clearly improve model performance compared to using the unity prior. Indeed, although the coupling-prior model predicts illusory repulsion when signals inform a shorter interval than the prior, the behavioral results confirmed general binding effects (i.e., compression) even in the 0 ms condition.

Model	P ($\xi = 1$)	μ_{AO}	ψ	P_{fr}	Log-likelihood	AIC
JPup	0.204	0	-	-	-131,281.18	262,566.36
MSup	0.786	0	-	-	-131,293.32	262,590.63
MAup	0.728	0	-	-	-131,359.65	262,723.31
PMup	0.693	0	-	-	-131,276.07	262,556.15
MIup	1	0	-	-	-260,633.01	521,268.03
JPcp	0.202	0	-	-	-131,296.44	262,598.88
MScp	0.769	0.119	-	-	-131,297.52	262,601.04
MAcp	0.759	0	-	-	-131,355.05	262,716.10
PMcp	0.702	0.046	-	-	-131,275.16	262,556.31
MIcp	1	271.642	-	-	-171,466.64	342,937.29
FC	-	-	198.837	-	-131,297.61	262,599.22
FR	-	-	-	0.018	-131,299.45	262,602.89
Null	-	-	-	-	-132,020.34	264,042.67

Table 1. Results of group-level fitting after excluding the observers whose responses followed the false-report model. Values in parentheses represent the fixed parameters. Hyphens indicate the parameters that are not included in the model. See the Fig. 2 caption for the meaning of the model names' acronyms.

These data suggest that, contrary to the assumption in the LT model, people tend to believe that their actions will immediately cause a consequence. Intentional binding, at least for an average observer, can be explained by unity integration as in classical multisensory models^{37,38}.

Nevertheless, setting the priors as free parameters is advantageous as it allows for a flexible description of binding tendencies. The current study confirmed substantial individual differences in the estimated parameters. Moreover, the causal and temporal priors depend on the task contexts. For example, the parameter for the causal prior should change by manipulating the action-outcome contingency^{54–56}. Additionally, people can learn the temporal structure of a task^{57–60}. Matute et al.⁶¹ also randomly presented 0 ms or 500 ms action-outcome intervals and revealed that the time estimate of the action shifted away from the outcome as more trials were experienced. One can reason that this was because the experience of delays led to larger time intervals being predicted. Consistent with this idea, when the interval length was blocked (not randomized), Glazebrook et al.⁶² observed strong binding for the 0 ms interval. It is thus possible that, in many studies that fix the interval to 250 ms, the prediction may approach that value (e.g., 230 ms), as Legaspi and Toyozumi suggested³⁶. However, importantly, the current analysis estimated the temporal prior at approximately zero, despite the presence of 250 and 500 ms conditions, indicating the robustness of the unity assumption as a prior⁶³.

According to the BCI models, the fact that intentional binding reflects both causal belief and temporal prediction may lead to difficulty in interpretation. Computational modeling can isolate these factors and explain binding tendencies in a certain individual and situation. It may help researchers investigate which (cognitive or sensory) processes are responsible for aberrant binding, for example, in schizophrenia^{40–42}. This approach may also be useful to explore the relationship between cognitive and sensory processes. While the current models assume dependency of temporal prior on causal inference, causal prior may also be subject to the temporal context. Although such influences could not be independently quantified in the current setting, future research can investigate their dynamics by manipulating causal and temporal contexts. Furthermore, contrary to the traditional analysis of binding, the computational models can provide trial-by-trial (probable) predictions of whether one would causally bind action and outcome. By assessing if this prediction would coincide with explicit causal judgments, researchers can explore a novel, powerful tool to investigate the sense of agency.

Possibility of non-Bayesian accounts

In principle, my findings support the idea that the BCI account can explain intentional binding. However, it is important to note that a substantial number of observers were better explained by a misunderstanding of the task demand, implying a potential problem with the typical measurements of intentional binding. This does not negate the occurrence of intentional binding in these observers but suggests that incidental errors might have inflated the observed binding effect. The classic clock task is usually designed to minimize such a risk by asking for action and outcome time reports in separate sessions. When the target events were mixed within a block, the binding effect was indeed enhanced⁶⁴. A critical limitation of traditional mean value analysis is its inability to differentiate between erroneous estimations and actual perceptual binding. As shown in this study, the computational modeling approach can help researchers manage such a risk.

Limitations

As a theoretical limitation of the present study, there may be other models that are better than those considered here. This study did not cover all possible accounts for intentional binding. Although I focused on models that assumed a common mechanism for perceptual shifts of keypress (action binding) and tone (outcome binding), some researchers have questioned this idea. For example, Tonn et al.⁶⁵ observed no correlation between the magnitude of action binding and that of outcome binding across individuals (see also²⁵). I also did not find similar changes in action and outcome binding as a function of interval length. This evidence raises the question of the validity of the overall compression as an index and suggests independent mechanisms for two subcomponents. Hon⁶⁶ provided a theoretical explanation for action and outcome binding, attributing them to attentional processes and predictive processes respectively. Nevertheless, to the best of my knowledge, there is currently no theory that independently provides quantitative predictions for each type of binding. Future studies are expected to build a computational model that explains the dissociation of the two binding types and compare it with the common mechanism model.

Furthermore, there are several practical limitations. First, the clock paradigm was used to measure intentional binding; however, other tasks, such as direct interval estimation^{18,19,21,22,67–70}, can also be used to measure this effect. Given that these tasks possibly engage different processes^{25,71}, the current findings may not be applicable to other measurements. Additionally, this study was conducted in an online setting, which may have introduced uncontrolled noise such as latency in tone presentation and variability in tone timing. These factors may have increased the length or uncertainty of the event interval. I believe the risk of these noises was minimized by manipulating the action-outcome intervals for an observer and incorporating individual baseline differences into the model. Nonetheless, it is worth noting that the absolute values of the estimated parameter (e.g., μ_{AO}) should be interpreted with caution. Moreover, the present study only quantified the mean, but not the variance, of the temporal prior distribution. It was difficult to estimate the variance while differentiating it from other parameters, such as $P(\xi = 1)$, in the current design³⁶. It is possible that there are also individual differences in the variance of this distribution, with some observers having a strong expectation for a specific interval length while others do not have such an expectation. Given that the perceived order of actions and outcomes is unlikely to be reversed, future models might consider asymmetric distributions like log-normal or gamma distributions for the temporal prior instead of the Gaussian distribution.

Conclusion

In conclusion, this study provides the first demonstration of fitting the computational models for intentional binding to trial-by-trial observation. The findings support the idea that people use BCI to recognize causal and temporal event structures. Additionally, probability-matching emerged as a possible candidate for implementing BCI. These results provide insight into the mechanisms underlying intentional binding and highlight the usefulness of computational modeling in understanding this phenomenon. Adopting this computational approach may also enable future research to characterize action-outcome binding patterns in both healthy and clinical populations by quantifying their causal and sensory priors.

Methods

Computational model

All the models shared the assumption regarding how ideal observers generate estimates in the baseline condition but differed in the assumptions in the operant condition. Here, I define the actual timing of an action as the onset time ($t_A^* = 0$). Although the tone occurs by itself in the baseline condition, I express the actual time of an outcome, t_O^* , by the corresponding action-outcome interval, t_{AO} (e.g., 250 ms), to compare with that in the operant condition ($t_O^* = t_A^* + t_{AO}$). The letters A and O in the parameter names represent 'action' and 'outcome,' respectively, while the asterisk denotes the actual value (physical timing). Our time perception is not precise in itself and is subject to internal noise and mechanical bias. The initial percept is thus expected to follow a Gaussian distribution whose mean is biased from the actual time and whose variance is the level of internal noises, namely,

$$\begin{aligned}\tau_A &\sim N(t_A^* + b_A, \sigma_A^2) \\ \tau_O &\sim N(t_O^* + b_O, \sigma_O^2)\end{aligned}\quad (1)$$

where b_A and b_O represent biases, σ_A and σ_O represent noises, and τ_A and τ_O represent event timing for action and outcome, respectively.

The task requires an *inverse inference* to find the most likely timing of action $\hat{t}_{baselineA}$ and outcome $\hat{t}_{baselineO}$ from the corresponding noisy signals, τ_A and τ_O , respectively. A Bayesian observer solves this by exploring the timing of action t_A and outcome t_O that maximize the likelihoods given by the conditional probabilities,

$$\begin{aligned}P(\tau_A|t_A) &= \frac{1}{\sqrt{2\pi}\sigma_A} \exp\left(-\frac{(\tau_A - t_A)^2}{2\sigma_A^2}\right) \\ P(\tau_O|t_O) &= \frac{1}{\sqrt{2\pi}\sigma_O} \exp\left(-\frac{(\tau_O - t_O)^2}{2\sigma_O^2}\right)\end{aligned}\quad (2)$$

As Eq. (2) takes on its maximum values when $\tau_A = t_A$ and $\tau_O = t_O$,

$$\begin{aligned}\hat{t}_{baselineA} &= \tau_A \\ \hat{t}_{baselineO} &= \tau_O\end{aligned}\quad (3)$$

Therefore, one can regard Eq. (1) as the distributions of baseline estimates and can obtain the approximation of the parameters, b_A , σ_A^2 , b_O , and σ_O^2 , from observation (see "Results").

Models based on Bayesian causal inference

In the operant condition, the observers experience two events, an action and an outcome. A Bayesian observer integrates these different signals by weighting their reliability to obtain the best estimates. Details can be found in Ref.³⁶; however, in summary, their BCI model further assumes that such integration occurs only when the observers believe the two events were correlated, that is, the action caused the outcome. The observer is thus challenged by two intertwined problems: when events occurred and whether they were causal. An optimal strategy maximizes the posterior probability, $P(t_A, t_O, \xi | \tau_A, \tau_O)$, that depends on the causal case, ξ . The binary parameter, ξ , indicates whether the action and outcome are relevant ($\xi = 1$) or not ($\xi = 0$). From the Bayesian theorem,

$$P(t_A, t_O, \xi | \tau_A, \tau_O) = \frac{P(\tau_A, \tau_O | t_A, t_O, \xi)P(t_A, t_O, \xi)}{P(\tau_A, \tau_O)}.\quad (4)$$

Given that the initial percepts, τ_A and τ_O , are independent of the causality ξ , $P(\tau_A, \tau_O | t_A, t_O, \xi) = P(\tau_A | t_A)P(\tau_O | t_O)$. Given $P(t_A, t_O, \xi)$ is a joint distribution of t_A , t_O , and ξ , $P(\tau_A, \tau_O | t_A, t_O, \xi)P(t_A, t_O, \xi) = P(\tau_A | t_A)P(\tau_O | t_O)P(t_A, t_O | \xi)P(\xi)$. Thus, the numerator that determines the peak location (i.e., maximum likelihood estimate) can be decomposed as follows:

$$P(\tau_A, \tau_O | t_A, t_O, \xi)P(t_A, t_O, \xi) = P(\tau_A | t_A)P(\tau_O | t_O)P(t_A, t_O | \xi)P(\xi).\quad (5)$$

The conditional probabilities, $P(\tau_A | t_A)$ and $P(\tau_O | t_O)$, are likelihoods of sensory signals given event timing and are given by Eq. (1). The $P(t_A, t_O | \xi)$, the prior of event timing, depends on the causality. When the action and outcome are correlated ($\xi = 1$), their gap should be normally distributed and, otherwise ($\xi = 0$), the prior distribution should be the uniform distribution, that is,

$$P(t_A, t_O | \xi) = \begin{cases} \frac{\exp\left(-\frac{(t_O - t_A - \mu_{AO})^2}{2\sigma_{AO}^2}\right)}{\sqrt{2\pi}\sigma_{AO}T} & (\xi = 1) \\ \frac{1}{T^2} & (\xi = 0) \end{cases}\quad (6)$$

where μ_{AO} is a prior for the action-outcome interval and σ_{AO} is its standard deviation. Moreover, T represents the finite integral ranges for t_A and t_O to normalize the probability distribution. Following³⁶, the fixed parameters, σ_{AO} and T , were set to 10 ms and 250 ms, respectively. Note that μ_{AO} is a free parameter in the model with the coupling prior, whereas it is fixed to zero in the model with the unity prior.

Combining Eqs. (2) and (6), an observer obtains the optimal estimates that maximize the posterior probability (i.e., the MAP estimates) in each causal scenario, Eq. (5). In the causal case,

$$\begin{aligned} \hat{t}_{A,\xi=1} &= \tau_A + \frac{\sigma_A^2}{\sigma_{tot}^2}(\tau_O - \tau_A - \mu_{AO}) \\ \hat{t}_{O,\xi=1} &= \tau_O - \frac{\sigma_O^2}{\sigma_{tot}^2}(\tau_O - \tau_A - \mu_{AO}) \end{aligned} \tag{7}$$

with the total variance, $\sigma_{tot}^2 = \sigma_A^2 + \sigma_O^2 + \sigma_{AO}^2$. In the acausal case, like the baseline estimates, the peak locations are computed simply based on $P(\tau_A|t_A)$ and $P(\tau_O|t_O)$:

$$\begin{aligned} \hat{t}_{A,\xi=0} &= \tau_A \\ \hat{t}_{O,\xi=0} &= \tau_O \end{aligned} \tag{8}$$

There are several possible ways to determine the final estimates using these conditional estimates. In the model proposed by Legaspi and Toyoizumi³⁶, one prefers the MAP of the joint posterior of causality and event timing. Because the estimates, t_A and t_O , at the peaks are known as Eqs. (7) and (8), the ratio of the peak values can be computed:

$$r \equiv \frac{\max_{t_A, t_O} P(t_A, t_O, \xi=1 | \tau_A, \tau_O)}{\max_{t_A, t_O} P(t_A, t_O, \xi=0 | \tau_A, \tau_O)} \tag{9}$$

The observer finally takes the estimates for the causal case with the higher peak (MAP);

$$\begin{aligned} \hat{t}_{operantA} &= \begin{cases} \hat{t}_{A,\xi=1} & \text{if } r > 1 \\ \hat{t}_{A,\xi=0} & \text{if } r \leq 1 \end{cases} \\ \hat{t}_{operantO} &= \begin{cases} \hat{t}_{O,\xi=1} & \text{if } r > 1 \\ \hat{t}_{O,\xi=0} & \text{if } r \leq 1 \end{cases} \end{aligned} \tag{10}$$

In addition, I implemented hierarchical strategies for causal inference. Instead of the MAP of the joint posterior, $P(t_A, t_O, \xi | \tau_A, \tau_O)$, the observer may use the marginal probability of each causal scenario, given by

$$\begin{aligned} P(\xi = 1 | \tau_A, \tau_O) &= \frac{P(\tau_A, \tau_O, \xi = 1)}{P(\tau_A, \tau_O, \xi = 1) + P(\tau_A, \tau_O, \xi = 0)} \\ P(\xi = 0 | \tau_A, \tau_O) &= 1 - P(\xi = 1 | \tau_A, \tau_O) \end{aligned} \tag{11}$$

with

$$\begin{aligned} P(\tau_A, \tau_O, \xi = 1) &= \int \int_R P(\xi = 1 | t_A, t_O, \tau_A, \tau_O) dt_A dt_O \\ &= \frac{P(\xi = 1) \sigma_{AO}}{\sqrt{2\pi} \sigma_{AO} T \sigma_{tot}} \exp\left(-\frac{(\tau_O - \tau_A - \mu_{AO})^2}{2\sigma_{tot}^2}\right) \end{aligned} \tag{12}$$

and

$$\begin{aligned} P(\tau_A, \tau_O, \xi = 0) &= \int \int_R P(\xi = 0 | t_A, t_O, \tau_A, \tau_O) dt_A dt_O \\ &= \frac{P(\xi = 0)}{T^2}. \end{aligned} \tag{13}$$

here, three strategies were considered to produce the final estimate. With the model-selection strategy, the observers consider only the most likely case:

$$\begin{aligned} \hat{t}_{operantA} &= \begin{cases} \hat{t}_{A,\xi=1} & \text{if } P(C = 1 | \tau_A, \tau_O) > 0.5 \\ \hat{t}_{A,\xi=0} & \text{if } P(C = 1 | \tau_A, \tau_O) \leq 0.5 \end{cases} \\ \hat{t}_{operantO} &= \begin{cases} \hat{t}_{O,\xi=1} & \text{if } P(C = 1 | \tau_A, \tau_O) > 0.5 \\ \hat{t}_{O,\xi=0} & \text{if } P(C = 1 | \tau_A, \tau_O) \leq 0.5 \end{cases} \end{aligned} \tag{14}$$

Meanwhile, with the probability-matching strategy, the use of estimates is stochastically determined in accordance with the probability of each case:

$$\hat{t}_{operantA} = \begin{cases} \hat{t}_{A,\xi=1} & \text{if } P(\xi = 1|\tau_A, \tau_O) > p \\ \text{where } p \in [0 : 1] & \text{uniform distribution} \\ \hat{t}_{A,\xi=0} & \text{if } P(\xi = 1|\tau_A, \tau_O) \leq p \\ \text{and sampled on each trial} \end{cases} \quad (15)$$

$$\hat{t}_{operantO} = \begin{cases} \hat{t}_{O,\xi=1} & \text{if } P(\xi = 1|\tau_A, \tau_O) > p \\ \text{where } p \in [0 : 1] & \text{uniform distribution} \\ \hat{t}_{O,\xi=0} & \text{if } P(\xi = 1|\tau_A, \tau_O) \leq p \\ \text{and sampled on each trial} \end{cases}$$

Finally, while the two strategies above exclusively employ either the causal or acausal case in a trial, the model-averaging strategy integrates estimates based on causal and acausal cases by weighting their likelihood:

$$\hat{t}_{operantA} = P(\xi = 1|\tau_A, \tau_O)\hat{t}_{A,\xi=1} + P(\xi = 0|\tau_A, \tau_O)\hat{t}_{A,\xi=0} \quad (16)$$

$$\hat{t}_{operantO} = P(\xi = 1|\tau_A, \tau_O)\hat{t}_{O,\xi=1} + P(\xi = 0|\tau_A, \tau_O)\hat{t}_{O,\xi=0}$$

Fixed-criterion model

As an alternative to the BCI model, the fixed-criterion model proposes that the observer simply averages the timings of action and outcome signals when their interval is below a certain threshold, ϕ :

$$\hat{t}_{operantA} = \begin{cases} \frac{\tau_A + \tau_O}{2} & \text{if } (\tau_O - \tau_A) \leq \phi \\ \tau_A & \text{if } (\tau_O - \tau_A) > \phi \end{cases} \quad (17)$$

$$\hat{t}_{operantO} = \begin{cases} \frac{\tau_A + \tau_O}{2} & \text{if } (\tau_O - \tau_A) \leq \phi \\ \tau_O & \text{if } (\tau_O - \tau_A) > \phi \end{cases}$$

False-report model

In the false-report model, the observer in the operant condition accidentally reports the timing of another event rather than the to-be-reported one at a certain probability, P_{fr} :

$$\hat{t}_{operantA} = \begin{cases} \tau_A & \text{if } P_{fr} > p \\ \text{where } p \in [0 : 1] & \text{uniform distribution} \\ \tau_O & \text{if } P_{fr} \leq p \\ \text{and sampled on each trial} \end{cases} \quad (18)$$

$$\hat{t}_{operantO} = \begin{cases} \tau_O & \text{if } P_{fr} > p \\ \text{where } p \in [0 : 1] & \text{uniform distribution} \\ \tau_A & \text{if } P_{fr} \leq p \\ \text{and sampled on each trial} \end{cases}$$

Null model

The baseline model assumes that the observers similarly estimate the timing in the baseline and operant conditions:

$$\hat{t}_{operantA} = \hat{t}_{baselineA} = \tau_A \quad (19)$$

$$\hat{t}_{operantO} = \hat{t}_{baselineO} = \tau_O,$$

predicting no intentional binding.

Experiment

Ethical considerations

This study was approved by the institutional review board of the University of Tokyo (no. 202119) and was conducted in accordance with the ethical standards of the 1964 Declaration of Helsinki. All observers provided informed consent prior to the commencement of the experiments.

Deviations from pre-registration

The experimental methods, including the sample size, procedures, and exclusion criteria were pre-registered (<https://aspredicted.org/rc8y6.pdf>). Although this included the analysis plan for summarized behavioral data, I changed the ANOVA design because I noticed that the registered one (2*3 repeated-measure ANOVA) was not consistent with the experimental design. The computational modeling analysis was exploratory. Only the rough plan was pre-registered in which I would compare models using the AIC by fitting the data at the group and individual levels and excluding participants whose responses would best be explained by either the false-report model or the null model.

Observers

I set 80 observers as the target sample size. According to Galang et al.⁴³, this sample size is sufficient to detect action and outcome shifts with typical effect sizes reported in a meta-analysis²⁵. Assuming that data from some

observers would be excluded from the analysis due to incomplete commitment and systematic issues, I invited 100 observers via an online crowdsourcing service (CrowdWorks; <https://crowdworks.jp/>). Eighty-three participants appropriately completed the experiment on time and I analyzed the data from 76 observers, excluding five left-handed ones and two with low performance (33 females; mean \pm SD of age = 41.28 ± 9.01).

Procedure

The experimental procedure was almost the same as that in the web-based experiment by Galang et al.⁴³, except that there were three levels of the action–outcome interval. The experiment was conducted on a web browser with a JavaScript application and a jsPsych plugin⁷². Observers viewed a clock rotating on a screen and estimated the timing of the target event based on the position of the clock hand. There were four types of tasks across 2 (measurement: baseline or operant) measurements by 2 (target event: keypress or tone) events. In the action baseline task, observers pressed the space bar and reported the timing of the keypress. In the outcome baseline task, they heard a tone externally generated at a random point between 1280 and 3840 ms from the trial onset and estimated its timing. In the operant action/outcome task, observers pressed a key and heard the tone followed by a certain interval, and they estimated the key/tone timing. The action–outcome interval was pseudo-randomly chosen from 0 ms, 250 ms, or 500 ms. Note that, considering the possibility that a longer duration of outcome tone introduces temporal variability, an auditory tone with a duration of 50 ms was used, instead of the 200 ms tone employed by Galang et al.⁴³. Observers performed different tasks in separate blocks. Each block of the baseline task contained 30 trials, whereas that of the operant task contained 90 trials comprising 30 repetitions of three possible action–outcome intervals. Observers completed two sets of randomly ordered task blocks (4 tasks \times 2 = 8 blocks), resulting in 60 trials per condition (i.e., 480 trials in total). Each block began with three practice trials, which were not analyzed.

Analysis

Behavioral data analysis

For each trial, an estimation error was calculated by subtracting the angle of the clock hand at the actual timing of the target event (keypress or tone) from the angle reported by the observer. These angle errors were transformed into temporal errors in ms. Positive (negative) numbers indicated overestimation (underestimation). For action and outcome timing, the mean errors were analyzed using one-way within-participant ANOVA with four conditions (baseline, 0 ms, 250 ms, and 500 ms operants). Moreover, I quantified the action and outcome binding by subtracting the average errors in the corresponding (key or tone) baseline errors from those in the operant conditions for each action–outcome interval length. Finally, intentional binding was calculated by summing up the amount of overestimation in action binding and underestimation in outcome binding. The binding effects were also analyzed using one-way ANOVA with three interval conditions. The p values below 0.05 were regarded as significant (two-tailed). Multiple comparisons were performed for significant main effects using Holm's method.

Parameter recovery

For all the models, I simulated estimates for the same number of observers and trials (i.e., 480 trials for 76 observers) as in my experiment. To-be-recovered parameters were sampled from the uniform distributions whose ranges [lower, upper] were set to [0, 1] for $P(\xi = 1)$, [0, 500] for μ_{AO} , [0, 1000] for ϕ , and [0, 1] for P_{fr} . I fitted each model to a simulated dataset and assessed Pearson's correlations between the sampled and recovered parameters.

Model recovery

In the model recovery process, I systematically fitted each of the 12 computational models to the datasets produced during the parameter recovery process, yielding a comprehensive 12 by 12 matrix of model fits. For each simulated observer within the dataset, the model yielding the lowest AIC score was identified as the best fitting model. The frequency with which each model emerged as the best fit was then calculated and expressed as a proportion of the total simulated observers within each set of simulated data.

Model fitting and comparison

All the models shared fixed parameters regarding baseline estimates, which were drawn from observation. The BCI models shared $P(\xi = 1)$ as a free parameter. Only the models with the coupling prior had an additional free parameter, μ_{AO} . The threshold and false-report models each had one free parameter, ϕ and P_{fr} , respectively. Finally, the baseline model was a null model without a free parameter. I fitted each model to the data from each observer at the individual level and to the pooled data at the group level. The fitting was performed using MLE, defining the best parameter values that maximized the likelihood of a given model providing the observed estimates. The likelihood was obtained by simulating the model by generating 100,000 pairs of τ_A and τ_O (i.e., Monte Carlo simulation). The Monte Carlo simulations give predictions about how responses would be distributed under a certain model for a given condition. Using this distribution, one can calculate how well the actual response in a given trial can be predicted (i.e., likelihood) by each model. I searched for the best parameters with R⁷³ and the R-package *pso*⁷⁴. The possible parameter ranges were the same as those for parameter recovery. I then compared the different models' abilities by reference to the best parameters based on the AIC, considering the difference in the parameter number. As a planned follow-up, I re-conducted the group-level fitting to the data excluding individuals whose responses were best accounted for by the false-report or baseline model.

Data availability

The datasets and R sources for analyses during the current study are available in the Open Science Framework repository at https://osf.io/zsh8g/?view_only=dc09ba21700441498af9e909a2cf900e. The pre-registration of the experiment can be found in the AsPredicted repository (#97173) at https://aspredicted.org/KMQ_MHC.

Received: 22 March 2023; Accepted: 27 January 2024

Published online: 05 February 2024

References

- Vallée-Tourangeau, F., Murphy, R. A. & Baker, A. G. Contiguity and the outcome density bias in action-outcome contingency judgements. *Q. J. Exp. Psychol. B* **58**, 177–192 (2005).
- Shanks, D. R., Pearson, S. M. & Dickinson, A. Temporal contiguity and the judgement of causality by human subjects. *Q. J. Exp. Psychol. B* **41**, 139–159 (1989).
- Bechlivanidis, C. & Lagnado, D. A. Does the ‘why’ tell us the ‘when?’. *Psychol. Sci.* **24**, 1563–1572 (2013).
- Blakey, E. *et al.* When causality shapes the experience of time: Evidence for temporal binding in young children. *Dev. Sci.* **22**, e12769 (2019).
- Buehner, M. J. & Humphreys, G. R. Causal binding of actions to their effects. *Psychol. Sci.* **20**, 1221–1228 (2009).
- Desantis, A., Waszak, F., Moutsopoulou, K. & Haggard, P. How action structures time: About the perceived temporal order of action and predicted outcomes. *Cognition* **146**, 100–109 (2016).
- Umamura, H. Causal context presented in subsequent event modifies the perceived timing of cause and effect. *Front. Psychol.* **8**, 314 (2017).
- Haggard, P., Clark, S. & Kalogeras, J. Voluntary action and conscious awareness. *Nat. Neurosci.* **5**, 382–385 (2002).
- Engbert, K. & Wohlschläger, A. Intentions and expectations in temporal binding. *Conscious. Cogn.* **16**, 255–264 (2007).
- Desantis, A., Roussel, C. & Waszak, F. On the influence of causal beliefs on the feeling of agency. *Conscious. Cogn.* **20**, 1211–1220 (2011).
- Moore, J. W. & Obhi, S. S. Intentional binding and the sense of agency: A review. *Conscious. Cogn.* **21**, 546–561 (2012).
- Buehner, M. J. Awareness of voluntary and involuntary causal actions and their outcomes. *Psychol. Consciousness Theory Res. Practice* **2**, 237–252 (2015).
- Dogge, M., Schaap, M., Custers, R., Wegner, D. M. & Aarts, H. When moving without volition: Implied self-causation enhances binding strength between involuntary actions and effects. *Conscious. Cogn.* **21**, 501–506 (2012).
- Kirsch, W., Kunde, W. & Herbold, O. Intentional binding is unrelated to action intention. *J. Exp. Psychol. Hum. Percept. Perform.* **45**, 378–385 (2019).
- Schwarz, K. A., Weller, L., Klaffehn, A. L. & Pfister, R. The effects of action choice on temporal binding, agency ratings, and their correlation. *Conscious. Cogn.* **75**, 102807 (2019).
- Tanaka, T. & Kawabata, H. Sense of agency is modulated by interactions between action choice, outcome valence, and predictability. *Curr. Psychol.* **40**, 1795–1806 (2021).
- Ruess, M., Thomaschke, R. & Kiesel, A. Intentional binding for unintended effects. *Timing Time Perception* **8**, 341–349 (2020).
- Buehner, M. J. Understanding the past, predicting the future: Causation, not intentional action, is the root of temporal binding. *Psychol. Sci.* **23**, 1490–1497 (2012).
- Suzuki, K., Lush, P., Seth, A. K. & Roseboom, W. Intentional Binding Without Intentional Action. *Psychol. Sci.* **30**, 842–853 (2019).
- Weller, L., Schwarz, K. A., Kunde, W. & Pfister, R. Something from nothing: Agency for deliberate nonactions. *Cognition* **196**, 104136 (2020).
- Poonian, S. K. & Cunnington, R. Intentional binding in self-made and observed actions. *Exp. Brain Res.* **229**, 419–427 (2013).
- Imaizumi, S. & Tanno, Y. Intentional binding coincides with explicit sense of agency. *Conscious. Cogn.* **67**, 1–15 (2019).
- Dewey, J. A. & Knoblich, G. Do implicit and explicit measures of the sense of agency measure the same thing?. *PLoS One* **9**, e110118 (2014).
- Hughes, G., Desantis, A. & Waszak, F. Mechanisms of intentional binding and sensory attenuation: The role of temporal prediction, temporal control, identity prediction, and motor prediction. *Psychol. Bull.* **139**, 133–151 (2013).
- Tanaka, T., Matsumoto, T., Hayashi, S., Takagi, S. & Kawabata, H. What makes action and outcome temporally close to each other: A systematic review and Meta-Analysis of temporal binding. *Timing Time Perception* **7**, 189–218 (2019).
- Klaffehn, A. L., Sellmann, F. B., Kirsch, W., Kunde, W. & Pfister, R. Temporal binding as multisensory integration: Manipulating perceptual certainty of actions and their effects. *Atten. Percept. Psychophys.* **83**, 3135–3145 (2021).
- Jagini, K. K. Temporal binding in multisensory and motor-sensory contexts: Toward a unified model. *Front. Hum. Neurosci.* **15**, 629437 (2021).
- Ernst, M. O. & Banks, M. S. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* **415**, 429–433 (2002).
- Lush, P. *et al.* Intentional binding as Bayesian cue combination: Testing predictions with trait individual differences. *J. Exp. Psychol. Hum. Percept. Perform.* **45**, 1206–1217 (2019).
- Yamamoto, K. Cue integration as a common mechanism for action and outcome bindings. *Cognition* **205**, 104423 (2020).
- Wolpe, N., Haggard, P., Siebner, H. R. & Rowe, J. B. Cue integration and the perception of action in intentional binding. *Exp. Brain Res.* **229**, 467–474 (2013).
- Beierholm, Shams & others. Comparing bayesian models for multisensory cue combination without mandatory integration. *Adv. Neural Inf. Process. Syst.* **20**, 1–8 (2007).
- Kayser, C. & Shams, L. Multisensory causal inference in the brain. *PLoS Biol.* **13**, e1002075 (2015).
- Shams, L. & Beierholm, U. R. Causal inference in perception. *Trends Cogn. Sci.* **14**, 425–432 (2010).
- Sato, Y., Toyoizumi, T. & Aihara, K. Bayesian inference explains perception of unity and ventriloquism aftereffect: Identification of common sources of audiovisual stimuli. *Neural Comput.* **19**, 3335–3355 (2007).
- Legaspi, R. & Toyoizumi, T. A Bayesian psychophysics model of sense of agency. *Nat. Commun.* **10**, 4250 (2019).
- Chen, Y.-C. & Spence, C. Assessing the role of the ‘unity assumption’ on multisensory integration: A review. *Front. Psychol.* **8**, 445 (2017).
- Ernst, M. O. Learning to integrate arbitrary signals from vision and touch. *J. Vis.* **7**(7), 1–14 (2007).
- Piray, P., Dezfouli, A., Heskes, T., Frank, M. J. & Daw, N. D. Hierarchical Bayesian inference for concurrent model fitting and comparison for group studies. *PLoS Comput. Biol.* **15**, e1007043 (2019).
- Haggard, P., Martin, F., Taylor-Clarke, M., Jeannerod, M. & Franck, N. Awareness of action in schizophrenia. *Neuroreport* **14**, 1081–1085 (2003).
- Moore, J. W., Dickinson, A. & Fletcher, P. C. Sense of agency, associative learning, and schizotypy. *Conscious. Cogn.* **20**, 792–800 (2011).
- Voss, M. *et al.* Altered awareness of action in schizophrenia: A specific deficit in predicting action consequences. *Brain* **133**, 3104–3112 (2010).

43. Galang, C. M., Malik, R., Kinley, I. & Obhi, S. S. Studying sense of agency online: Can intentional binding be observed in uncontrolled online settings?. *Conscious. Cogn.* **95**, 103217 (2021).
44. Loftus, G. R. & Masson, M. E. Using confidence intervals in within-subject designs. *Psychon. Bull. Rev.* **1**, 476–490 (1994).
45. Shams, L. & Beierholm, U. Bayesian causal inference: A unifying neuroscience theory. *Neurosci. Biobehav. Rev.* **137**, 104619 (2022).
46. Wozny, D. R., Beierholm, U. R. & Shams, L. Probability matching as a computational strategy used in perception. *PLoS Comput. Biol.* **6**, e1000871 (2010).
47. Fereday, R., Buehner, M. J. & Rushton, S. K. The role of time perception in temporal binding: Impaired temporal resolution in causal sequences. *Cognition* **193**, 104005 (2019).
48. Hilbe, J. M. *Negative Binomial Regression* (Cambridge University Press, 2011).
49. Gaissmaier, W. & Schooler, L. J. The smart potential behind probability matching. *Cognition* **109**, 416–422 (2008).
50. Mendonça, C., Mandelli, P. & Pulkki, V. Modeling the perception of audiovisual distance: Bayesian causal inference and other models. *PLoS One* **11**, e0165391 (2016).
51. Unturbe, J. & Corominas, J. Probability matching involves rule-generating ability: A neuropsychological mechanism dealing with probabilities. *Neuropsychology* **21**, 621–630 (2007).
52. Vul, E., Hanus, D. & Kanwisher, N. Attention as inference: Selection is probabilistic; responses are all-or-none samples. *J. Exp. Psychol. Gen.* **138**, 546–560 (2009).
53. West, R. F. & Stanovich, K. E. Is probability matching smart? Associations between probabilistic choices and cognitive ability. *Mem. Cognit.* **31**, 243–251 (2003).
54. Greville, W. J., Buehner, M. J. & Johansen, M. K. Causing time: Evaluating causal changes to the when rather than the whether of an outcome. *Mem. Cognit.* **48**, 200–211 (2020).
55. Shanks, D. R. Selective attribution and the judgment of causality. *Learn. Motiv.* **17**, 311–334 (1986).
56. White, P. A. Accounting for occurrences: A new view of the use of contingency information in causal judgment. *J. Exp. Psychol. Learn. Mem. Cogn.* **34**, 204–218 (2008).
57. Honda, T., Hirashima, M. & Nozaki, D. Habituation to feedback delay restores degraded visuomotor adaptation by altering both sensory prediction error and the sensitivity of adaptation to the error. *Front. Psychol.* **3**, 540 (2012).
58. Di Luca, M. & Rhodes, D. Optimal perceived timing: Integrating sensory information with dynamically updated expectations. *Sci. Rep.* **6**, 1–15 (2016).
59. Di Plinio, S., Arnò, S., Perrucci, M. G. & Ebisch, S. J. H. The evolving sense of agency: Context recency and quality modulate the interaction between prospective and retrospective processes. *Conscious. Cogn.* **80**, 102903 (2020).
60. Visalli, A., Capizzi, M., Ambrosini, E., Mazzonetto, I. & Vallesi, A. Bayesian modeling of temporal expectations in the human brain. *Neuroimage* **202**, 116097 (2019).
61. Matute, H., Cubillas, C. P. & Garaizar, P. Learning to infer the time of our actions and decisions from their consequences. *Conscious. Cogn.* **56**, 37–49 (2017).
62. Glazebrook, C. M., Gonzalez, C., Lyons, J. & Elliott, D. Temporal judgments of immediate and delayed consequences of self-initiated movements. *Can. J. Exp. Psychol.* **64**, 102–106 (2010).
63. Cravo, A. M., Haddad, H., Claessens, P. M. E. & Baldo, M. V. C. Bias and learning in temporal binding: Intervals between actions and outcomes are compressed by prior bias. *Conscious. Cogn.* **22**, 1174–1180 (2013).
64. Haggard, P. & Cole, J. Intention, attention and the temporal experience of action. *Conscious. Cogn.* **16**, 211–220 (2007).
65. Tonn, S., Pfister, R., Klaffehn, A. L., Weller, L. & Schwarz, K. A. Two faces of temporal binding: Action- and effect-binding are not correlated. *Conscious. Cogn.* **96**, 103219 (2021).
66. Hon, N. Attention and expectation likely underlie temporal binding measured using the Libet Clock. *Q. J. Exp. Psychol.* **76**, 2084–2093 (2023).
67. Caspar, E. A., Beyer, F., Cleeremans, A. & Haggard, P. The obedient mind and the volitional brain: A neural basis for preserved sense of agency and sense of responsibility under coercion. *PLoS One* **16**, e0258884 (2021).
68. Engbert, K., Wohlschläger, A. & Haggard, P. Who is causing what? The sense of agency is relational and efferent-triggered. *Cognition* **107**, 693–704 (2008).
69. Humphreys, G. R. & Buehner, M. J. Temporal binding of action and effect in interval reproduction. *Exp. Brain Res.* **203**, 465–470 (2010).
70. Ohata, R., Asai, T., Imaizumi, S. & Imamizu, H. I Hear My Voice; Therefore I Spoke: The Sense of Agency Over Speech Is Enhanced by Hearing One's Own Voice. *Psychol. Sci.* **33**, 1226–1239 (2022).
71. Siebertz, M. & Jansen, P. Diverging implicit measurement of sense of agency using interval estimation and libet clock. *Conscious. Cogn.* **99**, 103287 (2022).
72. de Leeuw, J. R. JsPsych: A JavaScript library for creating behavioral experiments in a web browser. *Behav. Res. Methods* **47**, 1–12 (2015).
73. R Core Team. *R: A Language and Environment for Statistical Computing*. (R Foundation for Statistical Computing, 2021).
74. Bendtsen, C. *Pso: Particle Swarm Optimization*. (2012).

Acknowledgements

We thank Editage (www.editage.com) for English-language editing.

Author contributions

The author confirms sole responsibility for the following: study conception and design, data collection, analysis and interpretation of results, and manuscript preparation.

Funding

This study was funded by Japan Society for the Promotion of Science (JP20K22269 and JP23K12928).

Competing interests

The author declares no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-53071-7>.

Correspondence and requests for materials should be addressed to T.T.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024