Temporal Binding in the Libet Clock Paradigm: A Psychometric and Mechanistic Point of View



Inaugural-Dissertation zur Erlangung der Doktorwürde der Fakultät für Humanwissenschaften der Universität Regensburg

vorgelegt von

Markus Siebertz

aus Bonn

- 2026 -

Regensburg, 2026

Gutachterin (Betreuerin): Prof. Dr. Petra Jansen

Gutachter: Prof. Dr. Gregor Volberg

Contents

Summary....1
1 Theoretical Background and State of Research....5
1.1 Sense of Agency....5
1.1.1 The Relevance of the Sense of Agency....5
1.1.2 Mechanistic Accounts of the Emergence of the Sense of Agency....6
1.2 Temporal Binding....8
1.2.1 The Beginning of Temporal Binding Research....8
1.2.2 Further Accounts of Processes Underlying Temporal Binding....9
1.2.3 Measuring Temporal Binding....10
1.2.4 Factors Influencing Temporal Binding....12
1.2.5 Expanding the Understanding of Temporal Binding Measures and Their Relationship. 14
1.3 Working Memory....16
1.3.1 The Possible Role of Spatial Working Memory in Measuring Temporal Binding with the Libet Clock Method....16
1.3.2 An Alternative Account of the Temporal Binding Phenomenon Based in Spatial Working Memory....18
2 Summary of the State of Research....19
3 Study 1: Diverging Implicit Measurement of Sense of Agency Using Interval Estimation and Libet Clock: Part I....22
3.1 Method....22
3.1.1 Sample....22
3.1.2 Material....23
3.1.3 Setting and Procedure....29
3.1.4 Study Design and Statistical Analysis....29
3.2 Results....32
3.2.1 Intentional Binding Measures....32
3.2.2 Correlations Between Binding Measures....35
3.3 Preliminary Discussion....38
4 Study 2: Diverging Implicit Measurement of Sense of Agency Using Interval Estimation and Libet Clock: Part II....40
4.1 Method....40
4.1.1 Sample....40
4.1.2 Material....41
4.1.3 Setting and Procedure....45
4.1.4 Study Design and Statistical Analysis....45

4.2 Results....46
4.2.1 Intentional Binding Measures....46
4.2.2 Correlations Between Binding Measures....50
4.3 Preliminary Discussion....52
4.3.1 Relationship Between Interval Estimation and the Libet Clock Paradigm....52
4.3.2 Reliability and Nomological Implications....52
4.3.3 Discussion of Exploratory Analyses....54
4.3.4 Limitations....55
4.3.5 Conclusion and Future Research....56
5 Study 3: Temporal Binding and Spatial Working Memory: A Correlational Approach....57
5.1 Method....58
5.1.1 Design....58
5.1.2 Sample....58
5.1.3 Material....59
5.1.4 Procedure....66
5.1.5 Statistical Analysis....66
5.2 Results....68
5.2.1 Action and Outcome Binding in the Libet Clock Task....68
5.2.2 Action and Outcome Binding in the Spatially Random Task....68
5.2.3 Relationship of Binding Between Binding Tasks....70
5.2.4 Relationship Between Binding and Working Memory Performance....70
5.3 Preliminary Discussion....73
6 Study 4: Temporal Binding and Working Memory: A Dual Task Approach....76
6.1 Method....77
6.1.1 Design....77
6.1.2 Sample....77
6.1.3 Material....78
6.1.4 Procedure....82
6.1.5 Statistical Analysis....83
6.2 Results....84
6.2.1 Memory Task Performance....84
6.2.2 Action Binding....85
6.2.3 Outcome Binding....87
6.2.4 Analysis of Three-Way-Interaction including Memory Type as a Factor....89
6.3 Preliminary Discussion....92

7 General Discussion....94
7.1 Integrative Overview....94
7.2 Theoretical implications....96
7.3 Practical implications....98
7.4 Limitations of this research....99
7.5 Suggestions for future research....101
7.6 Final conclusion....104
8 Declarations....105
8.1 Ethical Standards....105
8.2 Informed Consent to Participate and to Publish....105
8.3 Open Research Practices....105
8.4 Acknowledgements....105
8.5 Conflict of Interests....105
8.6 Funding....105
9 References....106

Summary

The sense of agency allows us to feel and recognize ourselves as the ones planning and executing our intentional actions (Moore, 2016). The phenomenon of intentional or causal binding has been proposed to be an objective measure of the sense of agency (Buehner, 2012; Haggard et al., 2002). This subjective shortening of the temporal interval between an action and its outcome has come to be called more neutrally temporal binding due to the ongoing discussion on its underlying mechanisms (e.g., Antusch et al., 2021; Buehner & Humphreys, 2009; Haggard et al., 2002).

Temporal binding has been investigated using various methods with the two most used paradigms being the Libet clock method and interval estimation. While these temporal binding paradigms have been used to draw conclusions regarding the sense of agency in various contexts (e.g., Aytemur & Levita, 2021; Dubynin et al., 2021; Pfister et al., 2014; Ricciardi et al., 2017) the role of intentional action in this phenomenon has been debated (Buehner, 2012; Kirsch et al., 2019; Suzuki et al., 2019). As mentioned before, there is an ongoing debate about the underlying mechanisms of temporal binding. The lack of knowledge of how and why exactly the interval between events in action sequences shrinks subjectively makes it difficult to draw conclusions regarding the sense of agency. It makes it especially hard to judge the appropriateness of temporal binding measures in various patient populations that might show deficits in processes that might contribute to temporal binding like abnormalities in time perception in schizophrenia (Lee et al., 2009). It also hinders the pooling of results and conclusions from studies that use different temporal binding paradigms since it is not known whether these paradigms measure the same processes. Additionally, there is little information on the psychometric properties of these paradigms.

We aim to contribute to filling this research gap with the four studies presented here. In Study 1 we employed a correlational within-participant design to compare temporal binding scores from two widely used paradigms, the Libet clock method and interval estimation. Both paradigms were completed by 46 healthy sports science students. Sample size was determined adaptively using Bayes

factors. Correlational analyses indicated no positive relationship between action or outcome binding in the Libet clock method and binding in the interval estimation paradigm. Neither did the results show a relationship between action and outcome binding within the Libet clock method. These results are backed by good to excellent reliability of temporal binding measures.

Study 2 repeats the design of Study 1 but aims to equalize experimental parameters between the two temporal binding paradigms which might have acted as confounding variables in Study 1, possibly occluding a positive relationship between the two. A variance in the action-outcome delay was introduced in the Libet clock method while the variance of action-outcome delays was reduced in the interval estimation paradigm to make both paradigms equal in this regard. Mirroring the interval estimation paradigm, passive conditions were included in the Libet clock method in addition to the usual active ones. We also preregistered Study 2. The adjusted paradigms were completed by 36 healthy sports science students with the sample size being determined via Bayes factors again. Despite these amendments, Study 2 could not show a positive relationship between the two methods.

Study 1 and Study 2 together show that temporal binding in the Libet clock method and in interval estimation likely do not share the same underlying processes. Starting from this result of the first two studies and concentrating on the Libet clock method, Study 3 and Study 4 aim to test a novel theory of how temporal binding in the Libet clock method might arise. The theory proposes that clock hand positions during target events (button presses or outcome tones) must be stored in spatial working memory to be reported later. In the contingent condition where both events occur and highlight neighboring clock hand positions, these positions might interfere pro- and retroactively with each other during the retention interval. This could make the stored representations more similar leading to the typical convergent shifts in reported clock hand position. Known characteristics of spatial working memory contribute to the plausibility of this theory. One such characteristic is similarity-based interference (D'Esposito & Postle, 2015) as the two clock hand positions are very similar stimuli. Another is the time course of encoding information from perception to working

memory that makes multiple to-be-encoded stimuli especially prone to interference when these stimuli are presented within a sub-second interval that is typically used in temporal binding paradigms (Klauer & Zhao, 2004).

To test this theory, Study 3 uses a preregistered correlational within-participant design. 104 healthy sport science students completed the Libet clock method and one visual and one spatial 2-back task. Visual working memory performance was tested to control for more general attentional compared to specific spatial working memory processes. Additionally, they completed a novel task similar to the Libet clock in which they were asked to report dot positions associated with button presses and outcome tones. This task was designed to separate the confound between the temporal and spatial dimensions of the Libet clock's continuously moving clock hand. We correlated scores from all four tasks with each other. We found action and outcome binding in the Libet clock task and outcome binding in the novel binding task. These were, however, not correlated. Outcome binding in the Libet clock showed a negative correlation to spatial working memory performance but evidence against a correlation with visual working memory performance was inconclusive according to Bayes factors. For outcome binding in the novel binding task the opposite was found. It correlated negatively with visual working memory performance but evidence regarding spatial working memory performance was inconclusive.

Study 4 aimed to compliment Study 3 with an experimental approach. We embedded each trial of the Libet clock task in the retention interval of a spatial (Corsi block task, Corsi, 1972) or visual (sequences of Chinese characters) memory task to put load on the specific memory modality. If spatial working memory processes form the basis of the binding effect in the Libet clock task, we expected to find increased binding with higher spatial memory load but not with higher visual memory load. 94 healthy sport science students completed the Libet clock task twice, embedded in the respective memory task. We used Bayesian mixed models to analyze our data. We also preregistered Study 4. The results suggest that action binding does not increase with spatial memory load, but the Bayes

factor indicates inconclusive evidence. Bayes factors show substantial evidence against an increase in action binding with higher visual memory load. Likewise, outcome binding does not seem to increase with spatial memory load. Again, Bayes factors remain inconclusive for this result but show conclusive evidence against an increase of outcome binding with higher visual memory load.

As mentioned above, Study 1 and 2 show that temporal binding in the Libet clock method seems to stem from different processes than temporal binding in the interval estimation paradigm and probably constitute separate phenomena. This highlights the question of what the respective underlying mechanisms behind temporal binding in various paradigms are. Study 3 and 4 propose and test a theory that tries to explain the binding phenomenon specifically for the Libet clock method. This theory views the binding effect as a result of spatial working memory processes. Testing this theory, the results of Study 3 and 4 do not confirm it. While we found a negative relationship between spatial working memory performance and outcome binding in the Libet clock (Study 3), we cannot exclude a similar relationship to visual working memory and thereby more general attentional processes. We were also unable to recreate the typical binding phenomenon when isolating the spatial aspect of the Libet clock method from the temporal one in the novel binding task. While we found evidence for outcome binding in the novel task, it was uncorrelated to its counterpart in the Libet clock paradigm. Spatial memory load did not seem to increase action or outcome binding in the Libet clock task. However, evidence was still inconclusive based on our data (Study 4). Despite our results speaking against an increase in any binding with higher visual memory load, the theory of binding being caused by spatial working memory processes is not supported by our data. We discuss implications of the results of the four studies presented here for theoretical considerations, practical application and future research.

1 Theoretical Background and State of Research[1]

1.1 Sense of Agency

The sense of agency enables us to perceive ourself as the originator of our actions and subsequent consequences (Moore & Obhi, 2012). It is necessary for a causal attribution of events to oneself, others or pure incidents in the environment. It also contributes to the recognition of oneself as an individual entity and the formation of a perceptible sense of self (Gallagher, 2000).

1.1.1 The Relevance of the Sense of Agency

The sense of agency is of special interest because of its role in the philosophical concept of free will (Aarts & van den Bos, 2011) in addition to its practical relevance: To influence one's environment in a goal-directed way the execution of intended actions must be monitored and their effects on the environment must be compared to the intended effects. Differentiating our actions and their consequences from those of others is of special importance in cooperative contexts pursuing a common goal.

Various psychiatric disorders and associated personality traits, including narcissism (Hascalovitz & Obhi, 2015), borderline personality disorder (Colle et al., 2020; Möller et al., 2020) and most notable schizophrenia (Di Plinio et al., 2019; Graham-Schmidt et al., 2016; Martin, 2013; Moore et al., 2011, 2013; Voss et al., 2010) show an altered sense of agency. The same is true for neurological disorders such as Parkinson's disease (Saito et al., 2017) and the anarchic and alien hand syndromes (Jenkinson et al., 2015; Moore & Fletcher, 2012).

[1] This section is in parts based on own previous published work: Siebertz, M., & Jansen, P. (2022). Diverging implicit measurement of sense of agency using interval estimation and Libet clock. *Consciousness and Cognition*, *99*, 103287. https://doi.org/10.1016/j.concog.2022.103287
and
Siebertz, M., & Jansen, P. (2025). Is the temporal binding effect in the Libet clock-task based in spatial working memory? A correlational and a dual-task approach. *Consciousness and Cognition*, 134, 103909. https://doi.org/10.1016/j.concog.2025.103909

A reduced he sense of agency also has to be considered in the legal context of criminal offenses under certain circumstances (De Pirro et al., 2020; Render & Jansen, 2020). Establishing a sense of agency in human-machine interaction (Cornelio Martinez et al., 2017; Endo et al., 2020; Sahaï et al., 2017) and motorized protheses (Schofield et al., 2019) prompts another challenge in this field.

1.1.2 Mechanistic Accounts of the Emergence of the Sense of Agency

Ongoing research regarding the underlying cognitive processes aims to explain the emergence of a sense of agency and to understand deficits of it in various conditions such as the ones mentioned above. A commonly used framework of motor control involves the combination of inverse and forward models to govern goal-directed motor behavior (Wolpert & Kawato, 1998). Inverse models derive a set of motor commands necessary to change a given environment toward a desired state. To act on the intention of picking up an object, an inverse model calculates the motor commands to move the hand above the object based on the relative positions and possible obstructions along the way. These motor commands are used to execute the intended action while a forward model simultaneously uses them and information about the current state of the environment to predict the outcome, that is the state of the environment after action execution. This internal prediction is compared to the actual perceptual feedback regarding the action effects. A mismatch results in an error signal which can be used to adjust the motor commands to account for the unexpected outcome and to further pursue the goal. When a sleeve gets caught at the corner of the table unexpectedly while reaching for an object, amendments to the original motor plan allow us to successfully continue our initial behavior. A match or mismatch between the predicted and perceived consequences in this comparator model is assumed to contribute or disrupt a sense of agency (Blakemore et al., 2002).

Inverse and forward models are also used to explain another phenomenon that might feed into the sense of agency: Sensory attenuation. Sensory attenuation consists of a diminished perceptual intensity of sensory feedback of voluntary actions and their effects in comparison to stimuli that are physically similar but not caused by oneself (Blakemore et al., 1999). It is measurable via event-related

sensory potentials in electroencephalography (EEG) (Gentsch & Schütz-Bosbach, 2011). Sensory attenuation is explained through the use of the prediction of the perception of action effects of the forward model in attenuating actual sensory feedback signals to a larger degree the smaller the mismatch error signal is (Blakemore et al., 1999). This leads to outcomes of voluntary, planned actions to be perceived as less intense. This might play a role in variations of the sense of agency.

These mechanisms establish a bottom-up emergence of the sense of agency. It is modelled as a product of low-level motor processes and the sensory input resulting from these processes. Research has also shown that the sense of agency is influenced by other, higher cognitive mechanisms. A top-down influence is exerted by beliefs about agency (Desantis et al., 2011, 2012; Lynn et al., 2014). In the account of optimal cue integration for the sense of agency (Synofzik et al., 2008a, 2008b, 2013) the sense of agency is differentiated into a sensorimotor feeling of agency which is closely related to low-level motor processes such as prediction and retrospective inference, and a cognitive judgement of agency which is based on the feeling of agency but also driven by beliefs about agency and information about the environment (e.g., what switches control which light sources in a room). These sensorimotor and cognitive levels also influence each other. For example, beliefs about the environment and control over it form motor predictions and experience from sensory input about action consequences inform our beliefs about our environment. Both levels are proposed to be subject to affective influences accounting for research showing a role of outcome valence and appraisal on the sense of agency (J. F. Christensen et al., 2019; Sarma & Srinivasan, 2021; Takahata et al., 2012; Tanaka & Kawabata, 2021; Yoshie & Haggard, 2017).

This account of optimal cue integration assumes that the formation of the sense of agency on both levels dynamically includes those informational sources most strongly that are regarded to be most reliable. An obviously related action-outcome sequence still might lead to a disrupted sense of agency if the outcome is manipulated to defy beliefs about the physical connection between action and outcome such as engaging a car's gas pedal while being in reverse gear by mistake.

Initially, methods to investigate the sense of agency consisted of probing participants for explicit agency judgements. One way to do this is by simply asking who caused an action or outcome (Farrer & Frith, 2002) or by asking participants to decide whether video feedback of hand movement shows their own or someone else's (Balslev et al., 2007). These explicit judgements can be biased as they are self-reports. Therefore, it is advantageous to additionally find implicit measures of the sense of agency.

1.2 Temporal Binding

One proposed implicit measure related to agency is the so-called intentional binding. This phenomenon shows in the convergent shift of two events in temporal perception. The events in question are a voluntary action and its consequence. They are subjectively shifted towards one another in time compared to involuntary actions or either event perceived independently (Haggard et al., 2002; Moore & Obhi, 2012). As justified below in the next section we will use the term temporal binding for this phenomenon to provide a neutral and consistent terminology. We deviate from this for discussions about this terminology. Early accounts of the connection between sense of agency and temporal binding relied on predictive motor processes and postdictive agency inference (Moore & Obhi, 2012) with following theories trying to reconcile the seemingly contradictory approaches (Synofzik et al., 2013).

1.2.1 The Beginning of Temporal Binding Research

Haggard et al. (2002) used the Libet clock (Libet et al., 1983) to measure temporal shifts during the perception of intentional actions and their consequences toward each other in their seminal study. Their procedure is described in more detail in the following section. These shifts were called intentional binding at first. They propose these shifts to be part of a cognitive process, binding actions and consequences together in order to process their causal dependency and voluntary nature. They also associate this cognitive process with motor control models. It is argued that an action intention is

necessary for this binding process, causing the formation of an efference copy which is used in a forward model to anticipate action outcomes in a comparator model (Blakemore et al., 2002).

1.2.2 Further Accounts of Processes Underlying Temporal Binding

Besides this mechanism based on prediction, another process discussed is retrospective inference (Wegner, 2003). The most likely cause for a perceived event is determined and is attributed to oneself if it aligns with a thought or intention. To test the roles of prediction and retrospective inference in binding, one study manipulated the known probability of the occurrence an outcome (Moore & Haggard, 2008). Binding was found to increase in trials when the outcome probability was high even when there was no actual outcome. This supports the role of predictive processes in binding. Binding also increased when the outcome probability was 50% making the outcome unpredictable but only if there was an actual outcome. This points to the role of retrospective inference in the context of binding.

Later research questioned the role of intention for the binding effect and explained previous findings with confounding variables that arise when trying to employ control conditions to isolate the effect of intention (Blakey et al., 2019; Buehner, 2012; Buehner & Humphreys, 2009; Fereday et al., 2019). Two studies employed either an apparatus that allowed passive movement of a button and the participant's finger (Kirsch et al., 2019) or virtual reality with haptic feedback (Suzuki et al., 2019) to closely match experimental conditions in which outcomes were triggered with and without voluntary action. Isolating the factor of intention, they found similar binding between active and passive conditions. Hoerl et al. (2020) argue that intention is only one of many cues to infer causality from and try to explain increased binding of intentionally caused events with an increase in causal attribution. This line of thinking led to the successively more general terms of causal binding and temporal binding. As the studies presented here investigate alternative mechanisms aside from intention and causation, we use the neutral term temporal binding from here on out.

Another approach to this assumes multisensory cue integration as the central process leading to temporal binding (Klaffehn et al., 2021; Lush et al., 2019; Wolpe et al., 2013; Yamamoto, 2020). Event sequences are perceived in various sensory and introspective modalities. For a coherent experience, a congruent representation must be formed including causal and temporal dependencies. Optimal multisensory cue integration proposes that perceived internal and external cues contribute more strongly to this representation the more reliable or precisely perceivable they are. Depending on their relative reliability certain events like an action or its consequence influence the representation of the event sequence in its temporal structure to a higher degree. This pulls or binds other events towards them.

Early work on temporal binding (Haggard et al., 2002) and current models of the sense of agency (Synofzik et al., 2013) view temporal binding as a perceptual phenomenon that can be used as a cue to derive agency over actions and outcomes. All other things being equal, stronger temporal binding should lead to an increased sense of agency, unless it is outweighed by other cues. In such cases, one might be able to infer a sense of agency from temporal binding although the strength of such inference is open to debate. This is further complicated by the notion that binding of the action and binding of the outcome stem from different mechanisms for which the link to the sense of agency has to be established independently (Tonn et al., 2021).

1.2.3 Measuring Temporal Binding

Different methods to measure temporal binding of actions and consequences have been conceived. Two common examples are the Libet clock paradigm and direct interval estimation of the action-outcome delay. Haggard et al. (2002) took the Libet clock (Libet et al., 1983) and asked participants for judgements of the clock hand position during their action, a voluntary key press, or its outcome, a sine wave tone. To compare these judgements, they employed control conditions in which the key press was substituted with passive finger movement induced by transcranial magnetic stimulation and baseline conditions in which only the key press or the tone occurred. Only voluntary key presses and

their outcomes were subjectively shifted towards each other in time. This method is the most common temporal binding measure. It also allows to split the binding effect into action binding and outcome binding. A more accessible variant designed for older people and children uses successively presented letters, a so-called stimulus stream, instead of the clock and participants name the letter visible during a target event (Cavazzana et al., 2014, 2015, 2017).

Direct estimation of the temporal interval is the second most commonly used method. It tries to avoid the effects of cross-modal temporal integration inherent to the Libet clock paradigm (Engbert et al., 2007). In the Libet clock paradigm, the visual, proprioceptive, and tactile sensory feedback, and related internal motor signals of the key press must be aligned temporally with the visual percept of the clock hand position. Interval estimations are often given numerically in milliseconds, either verbally or via keyboard. Interval reproduction is a variation of this. It circumvents the translation of the perceived interval into numbers (Humphreys & Buehner, 2010). A key is pressed for as long as the delay between two events lasted.

Additionally, a number of other methods were invented: Buehner and Humphreys (2009) wanted to prevent the influence of postdictive inference. They used a stimulus-anticipation task asking participants to press a key the moment they expected a tone to occur. Nolden et al. (2012) took a psychophysical approach with their method of constant stimuli. Participants compared constant action-outcome delays with tones of random duration. Other psychophysical methods use simultaneity judgements. They involve the detection of delays between actions and outcomes (Straube et al., 2017; van Kemenade et al., 2016; Villa et al., 2020; Wen et al., 2017) or between multiple outcomes of the same action (Desantis & Haggard, 2016; Eagleman, 2013). A related method lets participants report the order in which they perceive stimuli of which they are led to believe some to be self-generated and others to be generated by someone else (Haering & Kiesel, 2012). Even more indirectly, Kumar et al. (2015) compare time-dependent effects in attention capture and Vuorre and

Metcalfe (2017) compare alterations in visual motion illusions when the relevant stimuli are self- or other-elicited.

1.2.4 Factors Influencing Temporal Binding

Experimental results support the reasoning that temporal binding stems from a process that counts voluntary actions and their consequences specifically as self-generated and that this process fuels the sense of agency. These results show the influence of agency-related factors on temporal binding. They can be task-specific or traits of the participant. Our studies presented here employ and investigate the Libet clock method and interval estimation. Therefore, the following overview of task-related factors focusses on studies that use these two methods.

Freedom of action selection has been found to increase temporal binding in the Libet clock method (Barlas & Obhi, 2013) and in interval estimation (Barlas et al., 2017, 2018; Barlas & Kopp, 2018). However, Schwarz et al. (2019) could not replicate this finding. A similar increase of temporal binding was found for freedom of choice over the outcome. In the Libet clock paradigm, Beck et al. (2017) found increased temporal binding for action and outcome and Borhani et al. (2017) found increased temporal binding of the outcome. Sexual arousal (Render & Jansen, 2020), and fear and anger (J. F. Christensen et al., 2019) reduced temporal binding in the Libet clock paradigm. Both can be viewed as states of heightened arousal that can cause a loss of control. In interval estimation on the other hand, general arousal was found by Wen et al. (2015) to increase temporal binding.

Intentional actions are usually aimed at generating positive outcomes. Because of this, one line of research investigated the effects of outcome valence. In the Libet clock paradigm, Takahata et al. (2012) found that action outcomes associated with monetary loss lead to lower temporal binding. For predictable outcomes in the Libet clock paradigm, Yoshie and Haggard (2013) showed higher temporal binding for positive than for negative outcomes. J. F. Christensen et al. (2016), however, could not confirm this effect for neither, predictable and unpredictable outcomes. Similarly, Moreton

et al. (2017) did not find and effect of outcome valence. They used both interval estimation and the Libet clock method. Using consonant and dissonant piano chords as outcomes in an interval estimation task, Barlas et al. (2018) did not find an effect on temporal binding either. More methodological aspects of the task in use also effect temporal binding and its measurement. Forcing participants to delay action execution after they notice the intention to act increased temporal binding in the Libet clock paradigm (Vinding et al., 2013, 2015). While temporal binding has been shown to occur with action-outcome delays higher than one second in interval estimation (Humphreys & Buehner, 2009), binding effects appear to be strongest at around 250 ms of action-outcome delay (Ruess et al., 2017).

Intrapersonal factors seem to alter temporal binding, too, in addition to task-specific ones. A positive correlation between temporal binding and participants' self-reported beliefs in free will could be shown (Aarts & van den Bos, 2011). Results by Lynn et al. (2014) suggest that this might be a causal relationship. The other way around, falsely believing that an outcome was not self-generated reduce temporal binding (Desantis et al., 2011). Lush et al. (2016, 2019) found that trait differences in hypnotisability and mindfulness meditation experience influence temporal binding. They explain this by differences in the conscious perception of action intentions. Psychiatric disorders have also been found to be associated with alterations in temporal binding. This is found for schizophrenia (Graham-Schmidt et al., 2016; Voss et al., 2010), autism spectrum disorders (Zalla & Sperduti, 2015), borderline personality disorder (Möller et al., 2020) and obsessive-compulsive tendencies (Oren et al., 2019).

Summarized, the measurement of temporal binding with the Libet clock and interval estimation paradigms depends on task-specific and intrapersonal factors. Temporal binding has been used in a multitude of studies as an implicit measure of the sense of agency with a large methodological heterogeneity and partially contradictory results. However, the notion that temporal binding specifically measures processing of self-generated actions and their outcomes has been questioned.

1.2.5 Expanding the Understanding of Temporal Binding Measures and Their Relationship

Due to the methodological heterogeneity of temporal binding paradigms, it is important to assess what each of them measures and whether they measure the same. In this endeavor, it can be helpful to investigate the construct validity of measures pertaining to the sense of agency. Different paradigms measuring the same aspect of a person's pre-reflective sense of agency, e.g., temporal binding, should agree with each other in the sense of convergent validity. They should further agree with implicit measurement methods of different aspects of the pre-reflective sense of agency but to a lesser degree. This would show discriminant validity. Optimal cue integration (Synofzik et al., 2013) states that implicit measures represent parts of the underlying processes of the pre-reflective feeling of agency which feeds into conscious judgements of agency. This leads to the expectation that implicit measures coincide with explicit measures of the judgement of agency. An exception to this are situations where causality beliefs or other top-down cues override the information regarding the sense of agency provided by bottom-up cues such as temporal binding. To produce a better understanding of all of these constructs, a nomological network (Cronbach & Meehl, 1955) between implicit and explicit measures should be established by investigating their relationships.

Previous research covered parts of this nomological network. There are already some insights regarding relations between implicit measures, i.e., temporal binding and sensory attenuation. There are also insights regarding temporal binding and explicit agency judgements. As mentioned above, sensory attenuation describes the phenomenon of sensory feedback of self-generated action results being perceived as less intense compared to other-generated events (Blakemore et al., 1999). Borhani et al. (2017) report matching effects of how sensory attenuation and temporal binding react to experimental manipulations. Others studies come to opposing results (Poonian et al., 2015; Tsakiris & Haggard, 2003). Dewey and Knoblich (2014) found no significant relationship between temporal binding in interval estimation and separately measured sensory attenuation. Looking at existing research shows that different measures pertaining to the sense of agency are seldom measured in the

same study and are compared directly in even less studies. There are hints that temporal binding and sensory attenuation change similarly under some experimental manipulations. However, they do not seem to be related directly and therefore appear to measure different constructs. Hughes et al. (2013) reviewed potential processes underlying temporal binding and sensory attenuation and come to the conclusion that both are rooted in different mechanisms at least to some degree.

It is just as important to investigate how implicit and explicit measures relate to each other. Lynn et al. (2014) and Wen et al. (2015) found contrasting effects on both measures when manipulating free will beliefs and arousal, respectively. Contrarily, both measures reacted similarly to changes in action-outcome congruency (Ebert & Wegner, 2010). Some studies cannot find a direct correlation between implicit and explicit measures (Braun et al., 2014; Saito et al., 2015; Schwarz et al., 2019). An exception to this is a study by Imaizumi and Tanno (2019). While these studies come to mixed results in regard to the relationship between implicit and explicit measures, they also employ a large heterogeneity of methods for both levels of measurement.

As shown above, there is little research investigating the relationship between implicit measures and between implicit and explicit measures. Even less studies look at multiple measures of temporal binding at once. Haering and Kiesel (2014) applied violations of previously learned action-outcome associations and measured their effect on temporal binding. Using the method of constant stimuli, they found that a mismatch between learned and presented action outcome did not influence temporal binding. They confirmed this finding with interval estimation and the Libet clock paradigm. This experimental manipulation seems to have a similar effect on the different binding measures. However, this translates only partially to the question of how consistent different binding paradigms measure temporal binding on an intrapersonal level. A better understanding of this would facilitate the integration of studies with similar aims but different methods. For example, Suzuki et al. (2019) and Kirsch et al. (2019) investigated the role of intention in temporal binding but used only interval estimation or the Libet clock method. Their results can only be pooled with reservation.

1.3 Working Memory

Working memory is a set of cognitive functions that operate on memory content that is briefly represented and available for transformation and possible consolidation into long-term memory. Spatial working memory is a modality-specific subset of those processes that allow the transient retention and manipulation of spatial information. According to the multi-component working memory model (Baddeley, 2000) a central executive directs limited attentional resources towards specialized subsystems. Verbal memory content is stored in the phonological loop and visual information in the visuospatial sketchpad while the episodic buffer constitutes an interface between the two and additional information from long-term memory. The visuospatial sketchpad can be further divided into distinct working memory processes for objects defined by their visual features and for spatial information (Hartley et al., 2001; Klauer & Zhao, 2004; Smith et al., 1995). These processes share central attentional resources and can therefore be subject to interference (Vergauwe et al., 2009). They build an important foundation for a wide array of simple and complex tasks.

A basic function of spatial working memory is storing or maintaining spatial information. According to sensorimotor recruitment models, this happens through the continuous activation of systems that are involved in the perception of said information and attention-based rehearsal, meaning sequential assignment of attentional resources to the maintained location (D'Esposito & Postle, 2015). Another function is updating obsolete information. To prevent interference, information that is no longer relevant is removed from limited short-term storage and replaced by new information (Ecker et al., 2014; Miyake et al., 2000).

1.3.1 The Possible Role of Spatial Working Memory in Measuring Temporal Binding with the Libet Clock Method

At its core the Libet clock task is an effort of spatial working memory. A certain position or orientation of the clock hand indicated by the occurrence of an event, the action or the outcome, must be

remembered until the response phase at the end of the trial. During encoding and the retention interval this relevant spatial stimulus must be shielded from interference by all the irrelevant clock hand positions that are perceived before and after the target position as the clock hand constantly rotates. In the contingent conditions compared to the baseline, one additional position is highlighted by the irrelevant event. This position may be encoded involuntarily into spatial short-term memory. When the action is the target event, the later position during the outcome might interfere retroactively. When the outcome is the target event the earlier position during the action might interfere proactively. This interference would then shift the reported target position towards the interfering position.

Theoretically, the clock hand position during the irrelevant event is especially suitable to interfere with the one during the target event after the latter has been encoded. Similarity-based interference by superposition leads to a mixing of activation in the representational space between two stimuli that are held in working memory (Oberauer & Lin, 2017). According to the interference model of Oberauer and Lin, this representational overlap requires sufficient similarity between the interfering stimuli relative to the precision with which they are perceived. The clock hand positions during action and outcome in the Libet clock task with the often-used period of around 2500 ms and an offset of 250 ms are just around 36° apart.

Another possible point of interaction between the two positions is during the encoding and consolidation phase of transferring the perceived target position into short-term memory. Klauer and Zhao (2004) describe that during perception a sensory activation pattern is stored in a rapidly fading short-term memory. It must be consolidated into short-term memory using central resources to keep the representation available. The duration of this process is stated to be around 500 ms. During this time frame, into which the usual action-outcome delay in the Libet clock task falls, the representation of the target position might be especially susceptible to interference by another position that is highlighted by an attention-grabbing event. When the clock hand position during the action is the

target, the irrelevant position during the outcome is perceived while the former is still being encoded into short-term memory. When the position during the outcome is the target, attentional resources might still be bound by the consolidation of the position during the action. This fits the observation, that binding tends to decrease when the action-outcome delay moves from around 250 ms towards 500 ms and beyond (Ruess et al., 2017).

A similar argument can be drawn from the interference model of Oberauer and Lin (2017) who assign two functions to the so-called focus of attention in working memory. One is to hold a high-precision representation of a stimulus, and the other is to shield this representation from interference. They argue that the first function needs the focus of attention in working memory and in perception to coincide during encoding. Otherwise, the high resolution of the representation cannot be recovered retrospectively. A disparity of attentional focusses through the respective irrelevant event and associated clock hand position during encoding could be a potential mechanism behind the temporal binding effect. This also connects to the findings of Schwarz and Weller (2022), that attention-grabbing events other than the action or the outcome can cause binding in the Libet clock task.

1.3.2 An Alternative Account of the Temporal Binding Phenomenon Based in Spatial Working Memory

We propose an alternative explanation that binding in the Libet clock task is rooted in interference processes in spatial working memory. As shown above, working memory processes appear suited to cause or play a role in temporal binding. They fit in terms of the time scale and stimulus specificity. Naturally, spatial working memory processes cannot plausibly explain temporal binding effects in other paradigms such as interval estimation. To our best knowledge, this is a novel account and a connection between temporal binding and working memory in general and spatial working memory specifically has not been investigated before. We tested this alternative explanation of the emergence of the binding effect in Study 3 and Study 4. Initially, we employed a correlational approach in Study 3 and expanded on this with an experimental dual-task design in Study 4.

2 Summary of the State of Research

The sense of agency is a fundamental contributor to the human experience (Moore, 2016). Investigating it furthers the understanding of normal functioning and might contribute to a better understanding of conditions with alterations in the sense of agency like schizophrenia and inform possible interventions to help afflicted persons (Graham-Schmidt et al., 2016). A proposed way to access and objectively measure someone's sense of agency is via the phenomenon of temporal binding which is also called intentional or causal binding (Buehner, 2012; Haggard et al., 2002). This phenomenon is regarded to be observable in the subjective shortening of the temporal interval between an action and its outcome. Various accounts have been proposed to explain the emergence of this phenomenon. While some discuss the general mechanisms connecting volition and the perceived temporal shortening, others question the necessity of intentional action for the binding effect altogether (Haggard et al., 2002; Moore & Obhi, 2012; Schwarz & Weller, 2022; Suzuki et al., 2019). Further complicating this discussion is the multitude of paradigms that have been conceived to measure the temporal binding effect. The two most used methods, interval estimation and the Libet clock, utilize drastically different modalities to tap into participants' subjective experience. The Libet clock provides a visuo-temporal anchor which participants use to indicate their subjective perception. Interval estimation relies purely on one's introspective and reporting abilities. Until now there has been a lack of thorough methodological comparison of these methods. Specifically, the questions whether these methods capture the same underlying processes and whether they satisfy basic psychometric standards have been unaddressed.

We investigated the relationship of different measures of temporal binding, interval estimation and the Libet clock method, in Study 1 and Study 2. Study 1 adapted the two methods closely to how they were used in prior research and correlated binding scores from both methods across participants. We aimed to answer the question whether the two methods measure the same binding phenomenon with the same underlying mechanisms. To improve the interpretability of our

results, special interest was taken in the calculation of appropriate reliability scores. Using experimental parameters from prior research, the two methods differed in various characteristics. For example, the Libet clock method used a constant action-outcome delay of 250 ms while interval estimation used a variable delay. Following the recommendation of a reviewer during the publication process, Study 2 constitutes a follow-up experiment to Study 1 which controls and equalizes multiple of previously different experimental parameters. This allowed us to investigate the research question of the relationship between the two measures with greater certainty.

As mentioned before, there is an ongoing debate regarding the true nature of the processes behind the temporal binding phenomenon. Prominent accounts argue that motor control processes, specifically forward models, lead to temporal binding (Haggard et al., 2002). This puts an emphasis on the role of intentional action. Results from studies manipulating outcome characteristics like emotional valence or reward (e.g., Barlas et al., 2018) suggest that temporal binding is dependent on retrospective inference to at least some degree (Wegner, 2003). The necessity of intentional action for temporal binding to occur is further questioned by accounts that argue that a causal relationship between a cause and an outcome without intention on the side of the observing person suffices (Blakey et al., 2019; Buehner, 2012). Integrative accounts try to reconcile these seemingly contradictory approaches into comprehensive frameworks. This discussion can benefit from further research investigating possible mechanisms underlying temporal binding. Specifically, these accounts are partially conceived on a more phenomenological level and give no clear explanation of how the proposed mechanisms lead to the perception of a shortened action-outcome interval.

We contribute to this discussion by investigating the research question of whether spatial working memory processes are the cause of temporal binding in the Libet clock method in Study 3 and Study 4. Study 3 used a correlational design testing participants with the Libet clock to measure temporal binding, two 2-back tasks measuring visual and spatial working memory respectively and a binding task resembling the Libet clock method but replacing the element of the continuously moving

clock hand with directionally randomly appearing dots. Study 4 focused on the same research question but employed an experimental approach. The Libet clock method was sequentially embedded in the retention interval of a spatial and a visual working memory task varying working memory load by increasing the number of items to be remembered. The study's aim was to test whether specifically loading the spatial working memory binds resources that are needed to accurately perceive the Libet clock's action-outcome sequence and the visuo-temporal anchor of the moving clock hand. We assumed that binding these resources leads to a stronger binding phenomenon.

3 Study 1: Diverging Implicit Measurement of Sense of Agency Using Interval Estimation and Libet Clock: Part I[2]

This study aimed to contribute a first step in mapping the relations between different measures of temporal binding. We used variants of interval estimation and the Libet clock paradigm in the same participants. Temporal binding depends on a multitude of factors. Some are task-specific but many are relatively stable traits of the participants like free will beliefs (Aarts & van den Bos, 2011; Lynn et al., 2014), metacognitive access to intentions (Lush et al., 2016, 2019) and personality traits (Hascalovitz & Obhi, 2015; Oren et al., 2019). These should remain unchanged for consecutive paradigms in a single session. Both paradigms are regularly used to investigate differences in temporal binding temporal binding between healthy and clinical populations. This leads us to the hypothesis that temporal binding is positively correlated across participants between interval estimation and the Libet clock paradigm. We make no different predictions for action and outcome binding (Libet clock) and varying action-outcome delays (interval estimation).

3.1 Method

3.1.1 Sample

Sample size was not determined by power analysis. Instead, we used Bayes factors to determine when to terminate data collection. Initially, 48 Participants completed the experiment based on the final sample size of Suzuki et al. (2019). Of those, 10 were excluded for performing poorly in at least one of the two experimental tasks, hinting at a lack of proper task engagement (see the description of the tasks below for the respective exclusion criteria). Bayes factors indicated that this initial data did not provide enough evidence in favor or against some of the hypotheses to make a decision about accepting or rejecting them (see Section Study Design and Statistical Analysis). To gather more evidence for or against our hypotheses, 13 additional participants took part in the study. Of these, 3

[2] The results presented in this section were published in advance as Experiment 1 in: Siebertz, M., & Jansen, P. (2022). Diverging implicit measurement of sense of agency using interval estimation and Libet clock. *Consciousness and Cognition*, *99*, 103287. https://doi.org/10.1016/j.concog.2022.103287

were excluded due to poor performance. Three participants reported impaired and uncorrected sight or hearing and were excluded. One of these was already excluded due to low performance in both paradigms. One participant declared himself to have a neurological or psychiatric disorder. This participant was already excluded from analysis because of subpar performance in the interval estimation task leaving a final, valid sample of 46 participants. Bayes factors showed sufficient evidence in favor of or against the hypotheses and data acquisition was terminated. The sample consisted of healthy sports science students who were recruited via a newsletter used for communication between the department's teaching and student bodies. Participants received course credit. Of the sample included in the analysis, 36 (78.3 %) participants reported being female, 10 (21.7 %) male and none stated their gender as diverse. The mean age was 21.56 years (*SD* = 1.70, range = [18; 25]). 16 participants had participated in a study involving a Libet clock paradigm before. The study was executed in accordance with the declaration of Helsinki for the guidelines of ethical considerations. Ethical approval for this study was not required in accordance with the conditions outlined by the German Research Society (DFG) where research that carries no additional risk beyond daily activities does not require a research ethics board's approval. We communicated all considerations necessary to assess the question of ethical legitimacy of the study.

3.1.2 Material

3.1.2.1 Interval Estimation

The interval estimation paradigm was adapted from Suzuki et al. (2019) who conducted it more elaborately in virtual reality. It consisted of three blocks: an interval estimation practice block, and one block of an active and passive condition, respectively. During each trial of the practice block participants saw a fixation cross and a prompt reading 'ready' to indicate that they could press the spacebar which would trigger two pure sine wave tones at 880 Hz, each lasting 50 ms. The first tone was presented 200 ms after the key press. The interval between the tones was one of three base intervals (200, 400 or 800 ms) modified with a random jitter between -100 and +100 ms. Four trials

per base interval were presented (twelve practice trials total). 500 ms after the second tone the fixation cross and 'ready'-prompt disappeared. Participants entered their estimation of the interval between the tones using the number row of the keyboard. Participants confirmed their input with the enter key which caused the actual interval to be displayed in milliseconds. Following Suzuki et al. (2019), practice did not consist of estimating intervals between a key press and a tone to prevent improvements specific to active or passive button presses. Practice trials were not analyzed.

After the practice block the active and passive conditions followed with the order balanced across participants. These blocks were similar to the practice block with the difference that the key press triggered only one tone, and participants were asked to estimate the interval between the key press and the tone. The interval was one of the three base intervals (200, 400 or 800 ms) with a random jitter between -100 and +100 ms. 500 ms after the tone the prompt to enter their estimation was presented without feedback given. Per block, each base interval was presented for 17 trials (51 trials total per condition). In the active condition participants pressed the button themselves and in the passive condition participants observed the experimenter pressing the button. To eliminate confounding variables, the experimenter sat in a chair beside the participants in both conditions but remained inactive in the active one.

For both conditions a participant's mean judgement error was calculated by subtracting the actual from the estimated intervals. A positive judgement error indicates overestimation, and a negative judgement error indicates underestimation of the temporal intervals. The individual amount of temporal binding is then calculated by subtracting the mean judgement error of the active condition from the mean judgement error of the passive condition. This value is positive if on average the key press and the tone are perceived temporally closer in the active condition than in the passive condition, i.e., temporal binding is stronger when the button is pressed by the participants themselves compared to the button being pressed by the experimenter. Vice versa, if temporal binding is stronger

in the passive condition this value is negative, thus higher values indicate more temporal binding. Temporal binding was calculated overall and for each base interval separately.

As a measure of an individual participant's performance in the interval estimation task the Pearson correlation coefficient between actual and estimated intervals was calculated. Although participants varied widely regarding the scaling of their estimates, we reason that higher actual intervals should lead to higher estimated intervals, making a low correlation coefficient a suitable indicator of subpar performance. Due to the median's robustness against outliers, we chose correlation coefficients that are more than 1.5 times the interquartile range below the median as an exclusion criterion. Six participants failed to meet this criterion.

3.1.2.2 Libet Clock

The details of the Libet clock paradigm were adapted from the unpredictable session of the experiment by Kirsch et al. (2019) and are similar to the original paradigm by Haggard et al. (2002). Participants viewed a clock hand rotating clockwise (with a period of 2560 ms) on a clock face. The ticks on the clock face measured approximately 0.8° of visual angle in length, the circle had a diameter of 9.5° and the clock hand was 4.1° long.

Please refer to the letters A-D in Figure 1a) for the following description. The paradigm consisted of two contingent conditions (A and B) and two baseline conditions (C and D). All four conditions were presented in separate blocks with condition order being balanced across participants. In each trial of the contingent conditions the clock hand started rotating from a random position and participants were instructed to let it rotate at least once before freely pressing the spacebar at a time of their choice. They were also instructed to avoid pressing the spacebar at a predetermined or specific clock hand position. 250 ms after the key press a pure sine wave tone (1000 Hz, 100 ms duration) was presented. Before each block an instruction text appeared telling participants what to do in the upcoming block and whether they should pay attention to the clock hand position during the key press (contingent action condition, A) or during the tone (contingent outcome condition, B). The clock

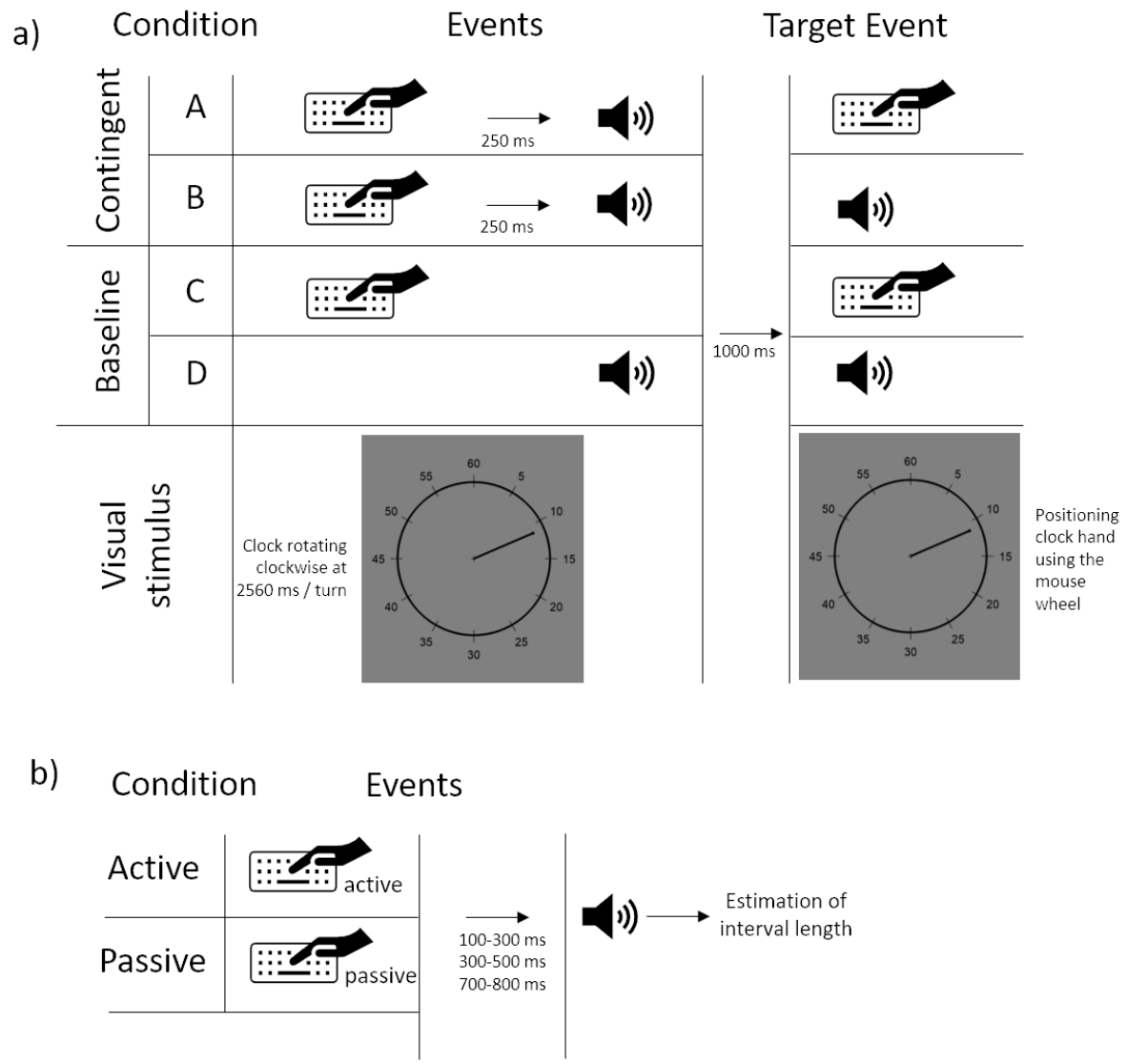
continued to rotate for a random amount of time between 1500 and 2500 ms after the key press or the tone, whichever was the target event. The clock then disappeared and after 1000 ms reappeared with the clock hand stationary in a random position between 45° and 60° clockwise or counterclockwise from the actual hand position during the target event. Participants used the mouse wheel to position the clock hand where it was during the target event and confirmed the position with the spacebar. Following this confirmation, the next trial started after 100 ms.

In the two baseline conditions only one event occurred. In the baseline action condition (C) participants pressed the key but no tone followed while in the baseline outcome condition (D) participants remained inactive and the tone occurred automatically 2810 to 4810 ms after the start of the trial representing one full rotation plus the 250 ms key press-tone offset plus 2000 ms of random jitter. As in the contingent conditions, participants subsequently reported the clock hand position during the occurring event.

A trial was repeated after a corresponding information text if participants (1) pressed the key before the clock hand made one full turn from where it started, (2) only pressed the button 10 seconds after trial begin, (3) did not press the button at all during the 12 seconds the clock was displayed rotating or (4) did not move the clock hand when reporting the perceived hand position. Participants practiced all conditions for five trials each after which they completed 36 trials per condition in the same condition order as during practice. Practice trials were not analyzed.

Figure 1

Trial schematic of the Libet clock paradigm (a) and interval estimation (b) in Experiment 1.



The spatial discrepancy between the actual and reported hand positions during the target event was converted into a temporal difference resulting in a judgement error. The judgement error was positive if the event was perceived as occurring later than it did, corresponding to a clockwise shifted hand position, and negative if the event was perceived as occurring earlier. For individual participants, trials were excluded from analysis if their judgement error exceeded the mean ± 3.5 times the standard deviation of their respective condition. In total 31 trials were excluded with a maximum of three trials per participant. A Participant's action binding was calculated by subtracting

their mean judgement error in the baseline action condition from their mean judgement error in the contingent action condition. Positive values mean that participants perceived the key press as occurring later in the contingent condition than in the baseline condition which would indicate action binding. Conversely, a participant's outcome binding is calculated by subtracting the mean judgement error in the contingent outcome condition from the mean judgement error in the baseline outcome condition. Here positive values mean that the tone was perceived as occurring earlier in the contingent condition than in the baseline condition which would show outcome binding. This deviates from previous research (e.g., Haggard et al., 2002) in order to obtain indices with higher values representing more binding for all types of binding. This is supposed to facilitate interpretability, as the aim of this study is to correlate temporal binding between the two paradigms. Total binding was calculated by adding action and outcome binding.

Some participants showed extremely early event reporting of contingent outcomes with mean judgement errors in the contingent outcome condition of up to -320 ms which equals a deviation of 45° from the actual hand position. Lush et al. (2019, supplementary material, https://osf.io/fnhgk/, retrieved: 22.07.2020) report similar values and argue that participants sometimes mistakenly report the timing of the key press instead of the tone in the contingent outcome condition. As the key press precedes the tone by 250ms this seemingly leads to very early reports of tone timing. Lush et al. handle this by excluding individual trials with outcome timing reports more than 200 ms before the event. Since participants receive clear instructions as to which event is to be judged ahead of every block and each condition is practiced beforehand, we argue that noticeably often judging the wrong event might indicate a lack of task engagement. To identify such participants, we used two-sample Kolmogorov–Smirnov tests to compare the empirical distribution of judgement errors in the contingent action condition with the empirical distribution of judgement errors in the contingent outcome condition after adding a constant of 250 ms to the latter. This tests the null hypothesis that the judgement errors in both conditions are sampled from the same underlying distribution or result from the same process,

i.e., the judgement of key press timing, considering location and shape of the distributions. To counteract the problem of multiple testing, which in this case would possibly lead to failing to identify problematic participants, we used Bonferroni correction. This method identified ten participants who seemed to have judged the wrong event. Three of those who did fell below the exclusion criterion for the interval estimation task as well. Altogether thirteen participants were excluded from analysis due to a lack in task performance in at least one paradigm.

3.1.3 Setting and Procedure

All stimuli were presented on a HP ProBook 650 G4 using PsychoPy (version 3.2.4, Peirce et al., 2019) under Windows 10. Participants observed visual stimuli on the notebooks display (39.6 cm diagonal, resolution: 1366x768, refresh rate: 60 Hz) at a viewing distance of approximately 60 cm and heard auditory stimuli via noise cancelling headphones. All participants' input and key presses were made using a standard external QWERTZ-keyboard with a key travel distance of 4 mm and the scroll wheel of a regular wired mouse. All participants gave written informed consent regarding the experiment and processing of their data. They then completed the two paradigms with the order being balanced across participants. At last, they completed a demographic questionnaire.

3.1.4 Study Design and Statistical Analysis

The aim of the present study was to investigate the coherence of naturally occurring, i.e., unaltered by experimental manipulation, temporal binding in the two paradigms. To this end the present study employs a cross-sectional correlational design without experimental manipulation. Measured and correlated variables encompass action, outcome, and total binding in the Libet clock paradigm, as well as binding in the interval estimation paradigm – overall and differentiated by base interval length.

All statistical analyses were performed in R (version 4.1.1, R Core Team, 2021). Pearson correlations, 95% CIs and associated Bayes factors between temporal binding measures from both paradigms were calculated. Additionally, odd-even split-half reliabilities for all types of binding were

calculated by splitting trials per condition and computing all binding types for each half of trials separately. After this the Spearman-Brown-corrected Pearson correlation coefficient for each binding type between the two halves was calculated.

Bayes factors trace back to the work of Jeffreys in the 1930's (Jeffreys, 1931; for a historical overview see Ly et al., 2016). When the prior probabilities of two hypotheses are equal, a Bayes factor is a likelihood ratio: the probability of the observed data under a null hypothesis compared to its probability under an alternative hypothesis. For equal likelihoods the Bayes factor is 1. Values going from 1 towards 0 indicate increasing evidence for one hypothesis while values above 1 indicate increasing evidence for the other (Dienes, 2014). A Bayes Factor *B* is directly interpretable as the data being *B* times more probable to have resulted under one hypothesis than under the other.

Besides other problems of significance-based hypothesis testing (summarized by Dienes, 2016), Bayes factors avoid the binary decision of rejecting or maintaining the null hypothesis. Instead, they either indicate sufficient evidence favoring one of the hypotheses over the other or insufficient evidence. At what value evidence is considered sufficient is an arbitrary choice and conventions have been established. Kass & Raferty (1995) argue that Bayes factors between 1/3 and 3 are indecisive. In other words, the data should be at least three times more likely under one hypothesis than under the other. We used this as a criterion for sample size, stopping data collection when all Bayes factors of interest were below 1/3 or above 3. Despite this cut-off criterion, the implied strength of evidence should be considered when interpreting the results. Bayes factors have been used in psychological studies and in studies regarding temporal binding in particular (e.g., Caspar et al., 2015; Haering & Kiesel, 2014; Lush et al., 2019; Suzuki et al., 2019).

We calculated Bayes factors for Pearson correlations between binding types adapting code originally published by Dienes (2008) and converted into R code by Baguley and Kaye (2010) as referenced on Dienes's website (Dienes, n.d.). The population value for H0 was set to zero. Following Dienes (2019), the prior distribution for modelling H1 was a positive half-normal distribution with

mode zero and a standard deviation equal to an expected effect size taken from relevant literature as a scaling factor. When no relevant effect size can be obtained from existing data, Dienes recommends finding a rough sensible maximum for the size of the effect. The expected effect is then set to half of that maximum. As we could not find a study that correlates binding resulting from different methods, we used data by Dewey and Knoblich (2014). They report correlations from an interval estimation paradigm between binding measures under different action-outcome delays. The highest of these within-method correlations was *r* = .72 between binding measures under 200 and 400 ms-delays. Correlations between binding measures under more different delays were lower. In our experiment the action-outcome delay was comparable between the two methods, leading us to use .72 as a rough sensible maximum for an expected effect of between-method binding correlation. This leads to a standard deviation for the model of H1 of $\sigma = \frac{Fisher's\ z(.72)}{2} = .454$.

We also calculated Bayes factors to see if binding was above zero on average. Again, the population value for H0 was set to zero and the model for H1 was a half normal distribution with mode zero and a scaling factor equal to an expected magnitude of binding. For expected effects for the means of the binding measures in the Libet clock paradigm we used data from Kirsch et al. (2019) and for interval estimation from Poonian et al. (2015). Mean binding occurs of around 18 ms for action binding, 40 ms for outcome binding and 57 ms for for total binding in the Libet clock paradigm. For interval estimation binding of around 26 ms is reported.

To check the dependence of our results on these priors we calculated robustness regions (RR). They indicate the range of values the scaling factors in our H1 models can take and still lead to the same qualitative conclusion regarding the hypotheses. The wider the RR the less sensitive a Bayes factor is to prior distribution assumptions. For correlations they are bound by 0 on the lower end with no upper limit, as we used one-tailed positive half-normal distributions as models for H1 with Fisher z-transformed correlations as scaling factors. As an artificial upper limit for RR computation, we used

a scaling factor of σ = 1.32 corresponding to a maximum expected effect of r = .99. This exceeds all plausible maximum expected effects for correlations between bindings. Similarly, for mean binding RRs are bounded by 0 on the lower end and have no upper limit. As a reasonable artificial upper limit, we used 10 times the expected effect to check for changes in resulting Bayes factors when varying the standard deviation of our H1-model. Following the reporting standards recommended by Dienes (2021), we complementarily report one sided *t*-tests, testing if correlations or means are above zero. As we base decisions regarding our hypotheses on Bayes factors, we set no significance level. According to Dienes (2021) however, Bayes factor thresholds of 3 and 1/3 are comparable in strictness to a significance level of 5%. Unlike a Bayes factor below the lower threshold (1/3 in this study), a p-value greater than 5% does not imply the null hypothesis to be true.

The experimental paradigm PsychoPy code, raw and congregated data, and R scripts producing all reported analyses are available (https://osf.io/gwjqf/). Note that the paradigms' instructions, feedback, etc. were presented in German but all text in the uploaded version has been translated into English.

3.2 Results

3.2.1 Intentional Binding Measures

Table 1 shows descriptive statistics and intercorrelations of the binding measures from the two paradigms in Study 1. For the Libet clock paradigm the average binding measures were 0.48 to 1.31 *SD* above zero. Bayes factors showed strong evidence for action binding being above zero ($t(45) = 3.28$, $p = .001$, $B_{HN(0, 18)} = 73.18$, $RR_{B>3}$ [2, >180]) and extreme evidence for outcome and total binding being above zero (outcome binding: $t(45) = 8.04$, $p < .001$, $B_{HN(0, 40)} >> 1000$, $RR_{B>3}$ [2, >400], action binding: $t(45) = 8.80$, $p < .001$, $B_{HN(0, 57)} >> 1000$, $RR_{B>3}$ [2, >570]). This means that temporal binding between action and outcome did occur, i.e., in the contingent conditions the temporal perception of action was shifted towards the outcome and vice versa. Outcome binding is more than five times higher than action

binding. RRs indicate no sensitivity to assumptions about prior distributions when using reasonable effect sizes. All three measures show high reliabilities (all rel ≥ .84).

Contrarily, Bayes factors in the interval estimation paradigm indicated insufficient evidence to decide between hypotheses for overall binding and under 200 and 800 ms base intervals (overall: $t(45) = 0.35$, $p = .366$, $B_{HN(0, 26)} = 0.41$, $RR_{1/3<B<3}$ [0, 33], 200 ms: $t(45) = 1.01$, $p = .160$, $B_{HN(0, 26)} = 0.65$, $RR_{1/3<B<3}$ [0, 53], 800 ms: $t(45) = 0.27$, $p = .394$, $B_{HN(0, 26)} = 0.60$, $RR_{1/3<B<3}$ [0, 53]). Under the 400 ms base interval moderate evidence for average binding being zero was found ($t(45) = -0.17$, $p = .567$, $B_{HN(0, 26)} = 0.31$, $RR_{B<1/3}$ [24, >260]). This means that the average judgement errors did not differ between the active and the passive condition under the 400 ms base interval. Although reliabilities are considerably lower than in the Libet clock paradigm, they all exceed .70. This matches consistent remarks of participants who found the conversion from perceived to numerical intervals to be difficult.

Table 1

Results of Experiment 1: Descriptive statistics for binding measures, their Pearson intercorrelations with 95% CI in brackets

Variable	*M* (*SD*)	Rel	1	2	3	4	5	6
Libet clock								
1. Action	13 (27)	.84	-					
2. Outcome	74 (63)	.96	-.04 [-.33, .25]	-				
3. Total	88 (67)	.96	.37 [.09, .59]	.92 [.85, .95]	-			
Interval Estimation								
4. Overall	3 (58)	.74	-.33 [-.56, -.04]	-.19 [-.45, .11]	-.31 [-.55, -.02]	-		
5. 200-ms BI	6 (44)	.75	-.22 [-.48, .07]	-.08 [-.36, .22]	-.16 [-.43, .13]	.65 [.44, .79]	-	
6. 400-ms BI	-2 (65)	.72	-.28 [-.53, .11]	-.18 [-.45, .12]	-.28 [-.53, .01]	.88 [.79, .93]	.59 [.36, .75]	-
7. 800-ms BI	4 (99)	.80	-.29 [-.54, .00]	-.17 [-.44, .12]	-.28 [-.53, .01]	.88 [.80, .93]	.30 [.01, .55]	.62 [.40, .77]

Note. Reliability was calculated as odd-even split-half reliability. BI = base interval.
n = 46.
M and *SD* are stated in milliseconds.

3.2.2 Correlations Between Binding Measures

For all correlations of binding measures between the two paradigms (shown in Table 1 and visualized in Figure 2). Bayes factors showed moderate evidence for the null hypothesis that the true correlation was zero over the alternative hypothesis that the true correlation lay above zero. This holds true for correlations of binding in the Libet clock paradigm with overall binding in interval estimation (action binding: $t(44) = -2.30$, $p = .987$, $B_{HN(0,\ 0.454)} = 0.10$, $RR_{B<1/3}$ [0.128, >1.32], outcome binding: $t(44) = -1.26$, $p = .892$, $B_{HN(0,\ 0.454)} = 0.15$, $RR_{B<1/3}$ [0.194, >1.32], total binding: $t(44) = -2.13$, $p = .981$, $B_{HN(0,\ 0.454)} = 0.11$, $RR_{B<1/3}$ [0.133, >1.32]), with binding in the 200 ms base interval (action binding: $t(44) = -1.50$, $p = .930$, $B_{HN(0,\ 0.454)} = 0.14$, $RR_{B<1/3}$ [0.171, >1.32], outcome binding: $t(44) = -0.53$, $p = .702$, $B_{HN(0,\ 0.454)} = 0.22$, $RR_{B<1/3}$ [0.295, >1.32], total binding: $t(44) = -1.10$, $p = .862$, $B_{HN(0,\ 0.454)} = 0.16$, $RR_{B<1/3}$ [0.212, >1.32]), with binding in the 400 ms base interval (action binding: $t(44) = -1.93$, $p = .970$, $B_{HN(0,\ 0.454)} = 0.12$, $RR_{B<1/3}$ [0.144, >1.32], outcome binding: $t(44) = -1.93$, $p = .884$, $B_{HN(0,\ 0.454)} = 0.16$, $RR_{B<1/3}$ [0.200, >1.32], total binding: $t(44) = -1.94$, $p = .970$, $B_{HN(0,\ 0.454)} = 0.12$, $RR_{B<1/3}$ [0.144, >1.32]) and with binding in the 800 ms base interval (action binding: $t(44) = -2.02$, $p = .975$, $B_{HN(0,\ 0.454)} = 0.11$, $RR_{B<1/3}$ [0.138, >1.32], outcome binding: $t(44) = -1.16$, $p = .873$, $B_{HN(0,\ 0.454)} = 0.16$, $RR_{B<1/3}$ [0.206, >1.32], total binding: $t(44) = -1.91$, $p = .969$, $B_{HN(0,\ 0.454)} = 0.12$, $RR_{B<1/3}$ [0.144, >1.32]). Depending on the correlation pair, RRs show that evidence would turn inconclusive when using scaling factors below 0.128 for the widest RR and 0.295 for the narrowest, corresponding to assumed rough maximum correlations of $r = .25$ and $r = .53$ respectively.

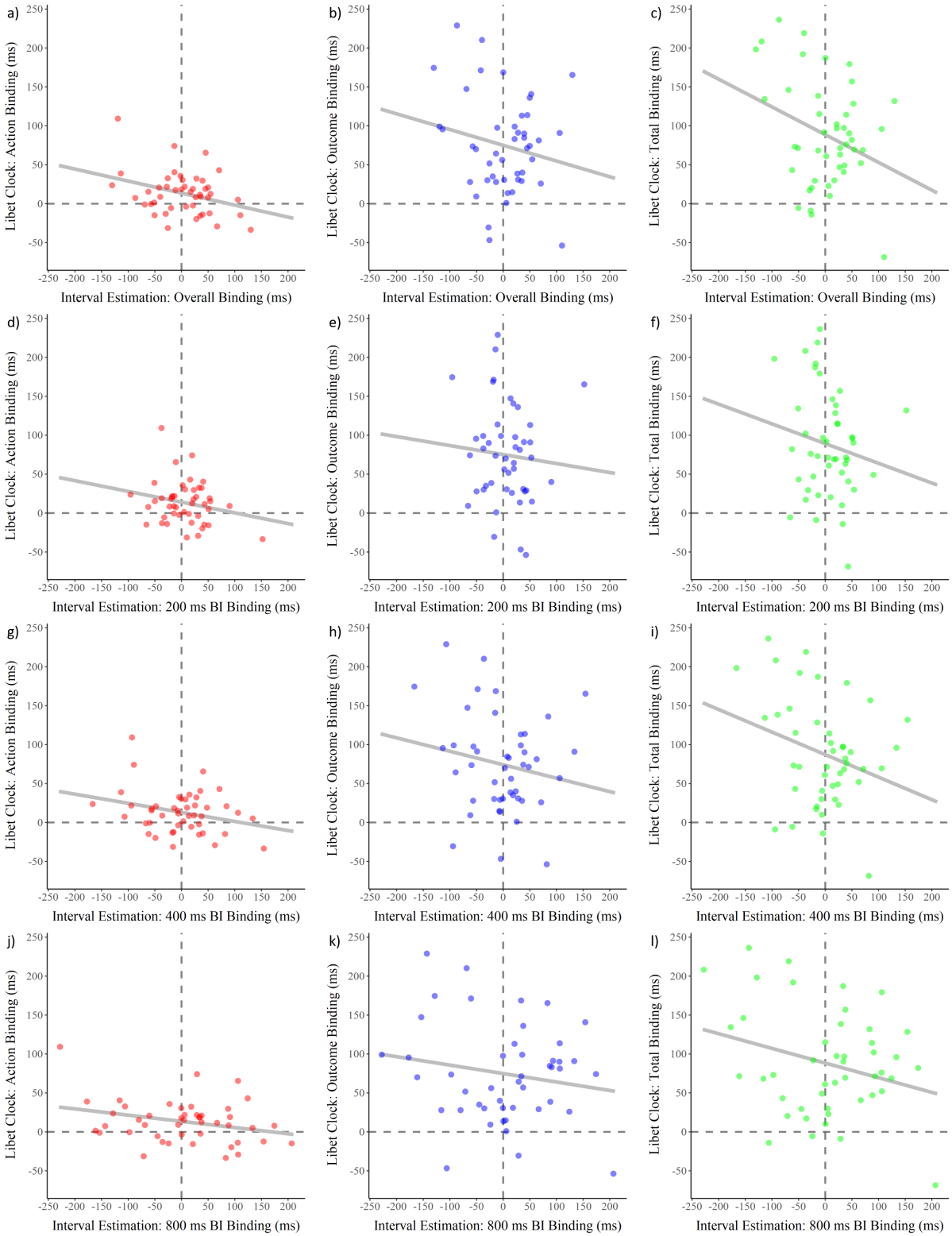
Study 1 has not been preregistered and the exclusion criteria for the Libet clock paradigm are somewhat novel. Therefore, we report the correlations between paradigms for which the decision based on the Bayes factors changes when including the participants who supposedly rated the wrong event in the contingent outcome condition. This applies to five correlations: Those of overall binding in interval estimation with outcome binding ($r = .11$, 95% CI [-.16, .37], $t(51) = 0.80$, $p = .214$, $B_{HN(0,\ 0.454)} = 0.61$, $RR_{1/3<B<3}$ [0, 0.869]) and with total binding in the Libet clock paradigm ($r = .05$, 95% CI [-.22, .32], $t(51) = 0.36$, $p = .361$, $B_{HN(0,\ 0.454)} = 0.40$, $RR_{1/3<B<3}$ [0, 0.549]); the correlation of binding in the 400 ms base

interval with outcome (r = .08, 95% CI [-.19, .35], $t(51) = 0.63$, p = .267, $B_{HN(0, 0.454)} = 0.51$, $RR_{1/3<B<3}$ [0, 0.711]) and with total binding in the Libet clock paradigm (r = .05, 95% CI [-.22, .32], $t(51) = 0.36$, p = .360, $B_{HN(0, 0.454)} = 0.40$, $RR_{1/3<B<3}$ [0, 0.549]); and one correlation of binding in the 800 ms base interval with total binding in the Libet clock paradigm (r = .08, 95% CI [-.19, .35], $t(51) = 0.60$, p = .277, $B_{HN(0, 0.454)} = 0.50$, $RR_{1/3<B<3}$ [0, 0.688]).

In exploratory analyses we investigate correlations within paradigms (see Table 1). Regarding those, moderate evidence for a lack of correlation between action and outcome binding in the Libet clock paradigm was found ($t(44) = -0.26$, p = .601, $B_{HN(0, 0.454)} = 0.27$, $RR_{B<1/3}$ [0.354, >1.32]), which would have been insufficient evidence for priors with a scale factor below 0.354 corresponding to a plausible maximum correlation below r = .61. In the interval estimation paradigm extreme evidence was found for positive correlations between binding in trials with base intervals of 400 ms and 200 ms ($t(44) = 4.86$, $p < .001$, $B_{HN(0, 0.454)} > 1000$, $RR_{B>3}$ [0.040, >1.32]), and 400 ms and 800 ms ($t(44) = 5.26$, $p < .001$, $B_{HN(0, 0.454)} > 1000$, $RR_{B>3}$ [0.040, >1.32]), respectively. This result is virtually independent of prior assumptions as shown by wide RRs as only a scale factor below 0.040, corresponding to a maximum plausible correlation of r = .08, would have changed this decision between hypotheses. For a positive correlation of binding between trials with base intervals of 200 ms and 800 ms the Bayes factor gave only moderate evidence with the correlation coefficient being smaller than the other two ($t(44) = 2.11$, p = .020, $B_{HN(0, 0.454)} = 4.14$, $RR_{B>3}$ [0.117, 0.736]). This evidence remains conclusive for sensible prior assumptions, with only scaling factors corresponding to assumed maximum plausible correlations below r = .23 or above r = .90 leading to different decisions.

Figure 2

Scatter plots showing the relationship of binding between paradigms in Study 1. Displayed pairwise are action, outcome and total binding in the Libet clock paradigm with binding in interval estimation: overall binding (a-c), binding in the in 200 ms base interval (d-f), the 400 ms base interval (g-i) and the 800 ms base interval (j-l). Simple regression lines are provided.



3.3 Preliminary Discussion

We aimed to investigate whether the binding effect in interval estimation and the Libet clock paradigm are at least partly dependent on the same intrapersonal factors. If so, we expected binding in both measures to be positively correlated across participants. Our results favor the hypothesis of the correlation between bindings to be zero above the hypothesis that they are positively correlated, assuming a rough plausible maximum correlation of $r = .72$. This holds for all subcategories of binding that we measured.

As a side result, temporal binding in the interval estimation task correlated positively across base intervals (200, 400, 800 ms) with moderate to extreme evidence according to Bayes factors. Interestingly, temporal binding in neighboring base intervals (200 and 400 ms, 400 and 800 ms) was correlated higher than between the two more distant intervals (200 and 800 ms). Dewey & Knoblich (2014) found the same pattern although they used unjittered intervals of 200, 400 and 1200 ms, interval reproduction instead of estimation, and compared intervals between a participant's key press and a tone to intervals between two tones presented without action. Temporal binding might be influenced differently by intrapersonal factors depending on the action-outcome delay, but this must be verified by future research, even more so for other methods like the Libet clock paradigm.

Even though our results are straightforward, one must consider that intention is not the only factor that varied between both paradigms' conditions. In the interval estimation task, we compared a consistently contingent self-agency condition (actions were always followed by a tone) with a consistently contingent other-agency condition. Besides intention, participants lacked proprioceptive and haptic feedback related to action execution in the latter. As for visual differences, they saw the experimenter's finger press the key instead of their own. What both conditions had in common was an active agent. The Libet clock paradigm on the other hand contrasted operant action-outcome

conditions with onset judgements of individually occurring events. While these conditions matched perceptually, the participant's intention is confounded with the causal action-outcome relationship.

4 Study 2: Diverging Implicit Measurement of Sense of Agency Using Interval Estimation and Libet Clock: Part II[3]

One reviewer rightly pointed out that the absence of a positive correlation between the two paradigms in our first experiment does not necessarily mean that the binding in the two methods stems from distinct mechanisms. Although we adapted the paradigms from existing research, they differed in certain features as discussed in the previous section. Following the reviewer's advice, we conducted a second experiment. The aim was to isolate the method of subjective time measurement as the distinctive feature by matching the two paradigms in interval lengths, stimuli and the conditions that are compared to derive the binding measures. Study 2 was preregistered (https://osf.io/jt9us) and PsychoPy code, data and analysis scripts are available (https://osf.io/gwjqf/). Again, our hypothesis was that binding measures between the two paradigms are correlated positively.

4.1 Method

4.1.1 Sample

Written informed consent was obtained from all participants. Once more, the sample consisted of students of sports science who received course credit for participation. As in Study 1, we used Bayes factors to determine when to stop data collection. For details see the section on statistical analysis below. 45 participants were tested. Three participants stated to have an acute psychiatric or neurological disorder, one participant stated to have impaired and uncorrected hearing, and one to have impaired and uncorrected sight. Some of these participants reported multiple exclusion criteria leading to four participants to be excluded from further analysis because of these reasons. Of the

[3] The results presented in this section were published in advance as Experiment 2 in: Siebertz, M., & Jansen, P. (2022). Diverging implicit measurement of sense of agency using interval estimation and Libet clock. *Consciousness and Cognition*, *99*, 103287. https://doi.org/10.1016/j.concog.2022.103287

remaining 41 participants, two met the criteria for low performance in the interval estimation paradigm and four in the Libet clock paradigm. See the respective section on the two paradigms below for details on exclusion criteria. Five participants met exclusion criteria for at least one paradigm. Of the final sample of 36 participants, 15 (41.7%) stated their gender as female, 21 as male and none as diverse. The mean age was 22.8 years (*SD* = 2.5, range = [19, 31]). Two participated in a study using the Libet clock before. None participated in Study 1.

4.1.2 Material

4.1.2.1 Interval Estimation

The interval estimation paradigm was largely adapted from Study 1. Instead of three base intervals of 200, 400 and 800 ms (±100 ms) only action-outcome intervals between 200 and 400 ms were presented. To increase reliability, a new trial number per condition was calculated using the Spearman-Brown prediction formula. Using the lowest reliability in Study 1 (r = .72 at 17 trials) and an aspired reliability of at least .80, 27 trials were estimated to be needed. To match the trial number in the Libet clock paradigm (see below), this was increased to 28 trials per condition. In both conditions, active and passive, the action-outcome intervals were fixed to 28 values evenly spaced across the closed interval [200 ms, 400 ms]. Interval order was randomized for each condition and participant. Conditions were counterbalanced across participants. The tone characteristics in both paradigms were changed to a frequency of 1000 Hz and a duration of 50 ms. As in Study 1, judgement errors were calculated by subtracting the actual interval from the estimated interval. Then the mean judgement error in the active condition was subtracted from the mean judgement error in the passive condition to derive a measure of temporal binding.

Because many participants reported difficulties with the interval estimation paradigm in Study 1, we expanded the training preceding the active and passive conditions. Participants were informed that all intervals ranged between 200 and 400 ms. Then, triggered by the participant pressing the spacebar, two tones (1000 Hz, 50 ms duration) were played with 200 ms delay accompanied by a

prompt on screen informing about the interval length. This was repeated six times, alternating between 200 and 400 ms to give participants a feel for interval lengths. These example intervals were followed by a training task during which two tones were presented randomly with a delay of 200 or 400 ms (equal probabilities for both interval lengths). Participants had to decide which interval length had been presented, familiarizing themselves with the input method for interval estimates. After each trial the actual interval length was displayed. They moved on once they were correct five times in a row. After wrong answers participants could choose to hear the sample intervals again. The same procedure, example intervals and training task, followed once more, this time with an additional interval length of 300 ms. Eight consecutive correct answers were required to continue. The final training task presented ten trials with ten tone-to-tone interval lengths spaced equally across the closed interval [200 ms, 400 ms] in a random order. Participants estimated the interval length and received the actual interval length as feedback but were not required to make correct estimates. Training trials were not analyzed.

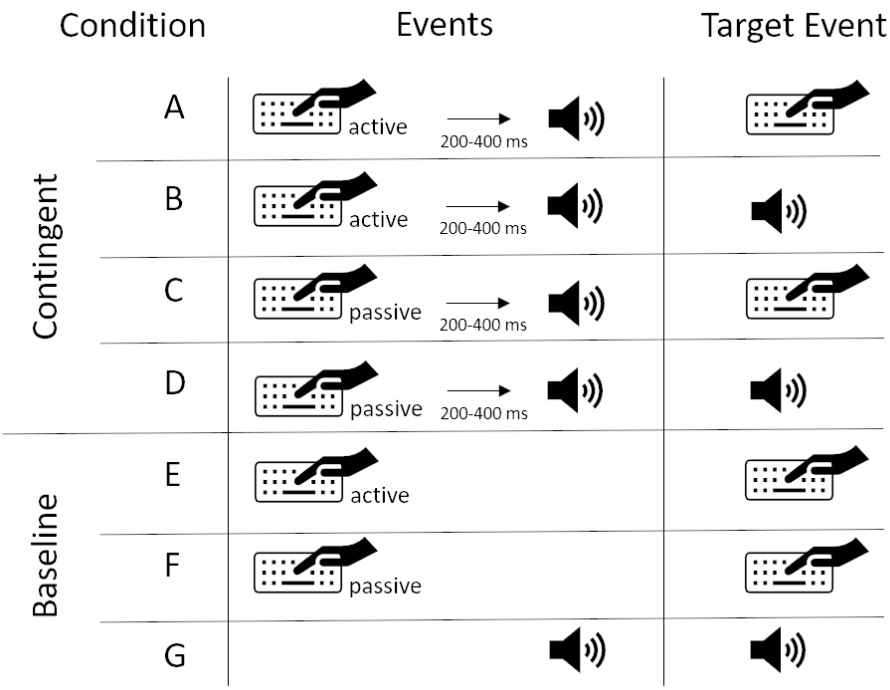
As a measure of performance, we calculated the correlation between estimated and actual intervals for each participant. Participants with a correlation more than 1.5 times the interquartile range below the median were excluded from further analysis.

4.1.2.2 Libet Clock

The procedure for the Libet clock paradigm was similar to the one in Study 1. To match the Libet clock paradigm to interval estimation in the way binding was calculated, we added three passive conditions to the usual four conditions employed in Study 1. Figure 3 shows all seven conditions. Please refer to the letters A-G in Figure 3 for the following description of binding measure calculation. Identically to the interval estimation paradigm, the three passive conditions (C, D and F) consisted of the experimenter pressing the button. Like in the contingent active conditions (A and B) which occurred as described for Study 1, participants estimated the clock hand position in the two contingent passive conditions (C and D) at the time of the button press or sound respectively. In the

Figure 3

Conditions in the Libet clock paradigm in Experiment 2.



passive baseline condition (F), the experimenter pressed the button without a following tone and the hand position at the time of the button press was estimated. The baseline tone condition (G) served as a baseline for both, the active and passive conditions in which the clock hand position at the time of the tone was to be estimated (B and D). To avoid sequence effects, 14 condition sequences were generated using a Latin square and counterbalanced across participants.

Action binding for the active and passive condition was calculated by subtracting the judgement error in the respective baseline condition (E and F) from the contingent condition (A and C). Outcome binding was calculated by subtracting the judgement error in the contingent condition (B and D) from the shared baseline condition (G). Adding action and outcome binding for the active and passive conditions yielded the respective total binding. To achieve a temporal binding measure like in the interval estimation paradigm, each binding type (action, outcome and total) in the passive

conditions was subtracted from its counterpart in the active conditions. Positive values indicate higher binding in the active compared to the passive conditions.

To reduce test length, we reduced the number of trials per condition, trying to retain reliabilities above .80. Using the Spearman-Brown prediction formula with the lowest reliability from Study 1 of r = .84 at 36 trials per condition led to 28 trials per condition for Study 2. Before the conditions, in sum consisting of 196 trials, participants did two practice trials per condition in the same condition order as during the main paradigm. These were not analyzed. Matching the interval lengths in the interval estimation paradigm, interval lengths in the Libet clock paradigm now were 28 evenly spaced values across the closed interval [200 ms, 400 ms] instead of the constant 250 ms interval length in Study 1. Like in the interval estimation paradigm, tone frequency was changed to 1000 Hz and tone duration to 50 ms.

Another change from Study 1 was to present an information screen between each trial reminding the participants of the current condition and the timing of which event, key press or tone, they were supposed to pay attention to. This was done because data from Study 1 suggest that many participants paid attention to the wrong event in some conditions. Additionally, the screen showed how many conditions and trials were left in the current condition (e.g., “condition 3 of 7” and “trial 15 of 28”, the latter being accompanied by a progress bar). The Libet clock paradigm can become monotonous and especially after adding three additional conditions we hoped to reduce the participants’ burden of losing track of the progress. To identify and exclude participants who mistakenly judged the action timing instead of tone timing in the contingent outcome conditions, we used the same method based on Kolmogorov–Smirnov test-based distribution comparisons of judgement errors as in Study 1. This was done for active and passive conditions separately. Being identified as supposedly having judged the wrong event in at least one of those led to exclusion from further analysis.

4.1.3 Setting and Procedure

All participants gave written informed consent regarding the experiment and processing of their data. They then completed the two paradigms with the order being counterbalanced across participants considering the condition sequence within paradigms. At the end, they completed a demographic questionnaire. During testing, the participant (on the right) and experimenter (on the left) were seated beside each other at a desk, separated by a transparent foil due to the ongoing COVID-19 pandemic at the time. In front of the participant stood a Lenovo TP E15 laptop (39.6 cm display diagonal, resolution: 1920x1080, refresh rate: 60 Hz, viewing distance approximately 60 cm) on which participants were presented all stimuli using PsychoPy (version 2021.2.2, Peirce et al., 2019) under Windows 10. Auditory stimuli were presented using the laptops built-in speakers, as headphones would have increased indirect contact between participants posing an infection risk. All participants' input and key presses were made using a standard external QWERTZ-keyboard in front of them with a key travel distance of 4 mm, and the scroll wheel of a regular wired mouse. The laptop's display was mirrored onto an external display in front of the experimenter. During passive conditions, the experimenter reached between the desk and the transparent foil leaving his hand in this position for the duration of the condition and pressed the spacebar when appropriate. After half of the testing session a five-minute break was used to air the room.

4.1.4 Study Design and Statistical Analysis

Study design and statistical analysis are largely the same as in Study 1. Again, measured and correlated variables encompass action, outcome and total binding in the Libet clock paradigm, as well as binding in the interval estimation paradigm. Due to the addition of passive conditions into the Libet clock paradigm, the three binding types are now calculated for active and passive conditions separately as well as for the difference of the two. In the interval estimation paradigm, no differentiation is made between interval lengths. We report correlations, 95% CIs and Bayes factors between binding

measures. Of interest for our hypothesis are the correlations between binding in the interval estimation paradigm on one hand and the various bindings in the Libet clock paradigm on the other.

We deviated from the preregistered analysis plan in three ways. First, for the sake of readability we report two instead of three decimal places. Second, we improved the calculation of the reliability of the binding measures. Using the same interval lengths in all conditions allowed us to calculate binding indices for each interval length separately. Initially we wanted to calculate the reliability estimator ω instead of the odd-even split half approach used in Study 1. Regrettably, our sample was too small for the required factor analytical approach. Instead, we randomly sampled 1000 combinations from all possible combinations of splitting the 28 intervals per binding type into two equal halves. We then calculated the Spearman-Brown corrected split-half reliability for each combination. We report the mean reliability complemented by the lowest and highest value after cutting off the lower and upper 2.5% of the 1000 values as limits for a 95% CI. Third, we additionally calculated Bayes factors to see if the binding measures in the Libet clock paradigm calculated as the difference between active and passive conditions average above zero with smaller expected effect sizes. The preregistered estimates were taken from Kirsch et al. (2019) and resulted from typical active conditions. When calculating binding values as the difference between such values and passive conditions, the effect sizes can be expected to be smaller. As smaller prior assumptions about effect sizes we used the 26 ms binding from Poonian et al. (2015), which resulted from an active-passive binding difference. We divide this into the ratio of action and outcome binding found by Kirsch et al. giving an expected effect size of 8 ms for action, 18 ms for outcome and 26 ms for total binding.

4.2 Results

4.2.1 Intentional Binding Measures

Table 2 shows descriptive statistics and intercorrelations of the binding measures from the two paradigms in Study 2. For the active conditions in the Libet clock paradigm corresponding to the measures in Study 1, average binding is approximately 0.3 to 0.9 *SD* above zero. Bayes factors show

moderate evidence for action binding above zero ($t(35) = 1.96$, $p = .029$, $B_{HN(0, 18)} = 3.87$, $RR_{B>3}$ [6, 29]) which becomes inconclusive for prior expected binding below 7 ms and above 29 ms. The evidence for outcome and total binding being above zero is extreme (outcome binding: $t(35) = 5.50$, $p < .001$, $B_{HN(0, 40)} >> 1000$, $RR_{B>3}$ [2, >400], total binding: $t(35) = 4.71$, $p < .001$, $B_{HN(0, 57)} >> 1000$, $RR_{B>3}$ [3, >570]) which holds for any sensible prior assumptions.

In the passive conditions average binding is around 0.1 to 0.7 *SD* above zero while the evidence regarding action binding is inconclusive ($t(35) = 0.80$, $p = .216$, $B_{HN(0, 18)} = 0.59$, $RR_{1/3<B<3}$ [0, 34]). Bayes factors for outcome and total binding show extreme evidence for the two to be above zero on average (outcome binding: $t(35) = 4.10$, $p < .001$, $B_{HN(0, 40)} > 1000$, $R_{RB>3}$ [2, >400], total binding: $t(35) = 3.89$, $p < .001$, $B_{HN(0, 57)} = 528.43$, $R_{RB>3}$ [3, >570]), holding for any sensible prior assumptions.

The binding measures calculated as the difference between active and passive conditions analogous to the interval estimation paradigm lie roughly 0.2 to 0.3 *SD* above zero. With the preregistered expected effect sizes of 18, 40 and 57 ms for action, outcome and total binding, respectively, Bayes factors yield inconclusive results (action binding: $t(35) = 1.21$, $p = .118$, $B_{HN(0, 18)} = 1.37$, $RR_{1/3<B<3}$ [0, 98], outcome binding: $t(35) = 1.59$, $p = .060$, $B_{HN(0, 40)} = 1.22$, $RR_{1/3<B<3}$ [0, 156], total binding: $t(35) = 1.87$, $p = .035$, $B_{HN(0, 57)} = 2.20$, $RR_{1/3<B<3}$ [36, 416]). Using prior effect assumptions of 8, 18 and 26 ms, respectively, results are still inconclusive for action and outcome binding (action binding: $t(35) = 1.21$, $p = .118$, $B_{HN(0, 8)} = 1.63$, $RR_{1/3<B<3}$ [0, >80], outcome binding: $t(35) = 1.59$, $p = .060$, $B_{HN(0, 18)} = 2.14$, $RR_{1/3<B<3}$ [0, 156]) but evidence for total binding on average being above zero becomes moderate ($t(35) = 1.87$, $p = .035$, $B_{HN(0, 26)} = 3.43$, $RR_{B>3}$ [12, 36]). The latter stays conclusive for a relatively small band of prior assumptions with 36 ms not seeming too unreasonable for an expected effect. The reliability of binding measures in the Libet clock paradigm are very high and even the lower bounds of the 95% CIs are acceptable with the exemption of action binding in the passive condition whose lower bound is just below .7.

The one binding measure in the interval estimation paradigm averages roughly 0.6 *SD* above zero. The Bayes factor supports this showing strong evidence ($t(35) = 3.81$, $p < .001$, $B_{HN(0, 26)} = 323.05$, $RR_{B>3}$ [1, >260]) which holds for all sensible priors. Despite increasing the number of trials per condition and training, reliability is not satisfactory.

Table 2

Results of Experiment 2: Descriptive statistics for binding measures, their Pearson intercorrelations with 95% CI in brackets

Variable		*M* (*SD*)	Rel [95% CI]	1	2	3	4	5	6	7	8	9
Libet clock												
Active	1. Action	15 (46)	.91 [.85, .96]	-								
	2. Outcome	46 (50)	.92 [.87, .95]	.31 [-.02, .58]	-							
	3. Total	61 (78)	.94 [.90, .96]	.79 [.62, .89]	.83 [.68, .91]	-						
Passive	4. Action	4 (32)	.79 [.67, .88]	.09 [-.25, .40]	.14 [-.20, .45]	.14 [-.20, .44]	-					
	5. Outcome	34 (49)	.87 [.80, .93]	.11 [-.23, .42]	.56 [.28, .75]	.42 [.11, .66]	-.01 [-.34, .32]	-				
	6. Total	38 (58)	.85 [.77, .91]	.14 [-.20, .45]	.55 [.26, .74]	.43 [.12, .67]	.54 [.25, .74]	.84 [.70, .91]	-			
Active-Passive	7. Action	11 (54)	.86 [.77, .92]	.80 [.65, .90]	.18 [-.16, .48]	.59 [.33, .77]	-.52 [-.73, -.23]	.10 [-.24, .41]	-.20 [-.50, .13]	-		
	8. Outcome	12 (47)	.85 [.75, .92]	.22 [-.12, .51]	.49 [.19, .70]	.44 [.13, .67]	.16 [-.18, .47]	-.45 [-.68, -.15]	-.29 [-.57, .04]	.09 [-.25, .41]	-	
	9. Total	23 (74)	.86 [.78, .92]	.72 [.51, .85]	.44 [.13, .67]	.71 [.49, .84]	-.28 [-.55, .06]	-.21 [-.51, .12]	-.33 [-.59, .00]	.78 [.61, .88]	.69 [.47, .83]	-
Interval Estimation												
	10. Overall	13 (21)	.63 [.44, .79]	-.13 [-.44, .21]	-.31 [-.58, .02]	-.28 [-.55, .06]	-.14 [-.45, .20]	-.16 [-.46, .18]	-.21 [-.50, .13]	-.03 [-.35, .30]	-.17 [-.47, .17]	-.12 [-.44, .21]

Note. Reliability was calculated as the mean of 1000 randomly drawn split-half reliabilities and the 95% CI marks the highest and lowest values excluding the 25 highest and lowest reliabilities. n = 36. *M* and *SD* are stated in milliseconds.

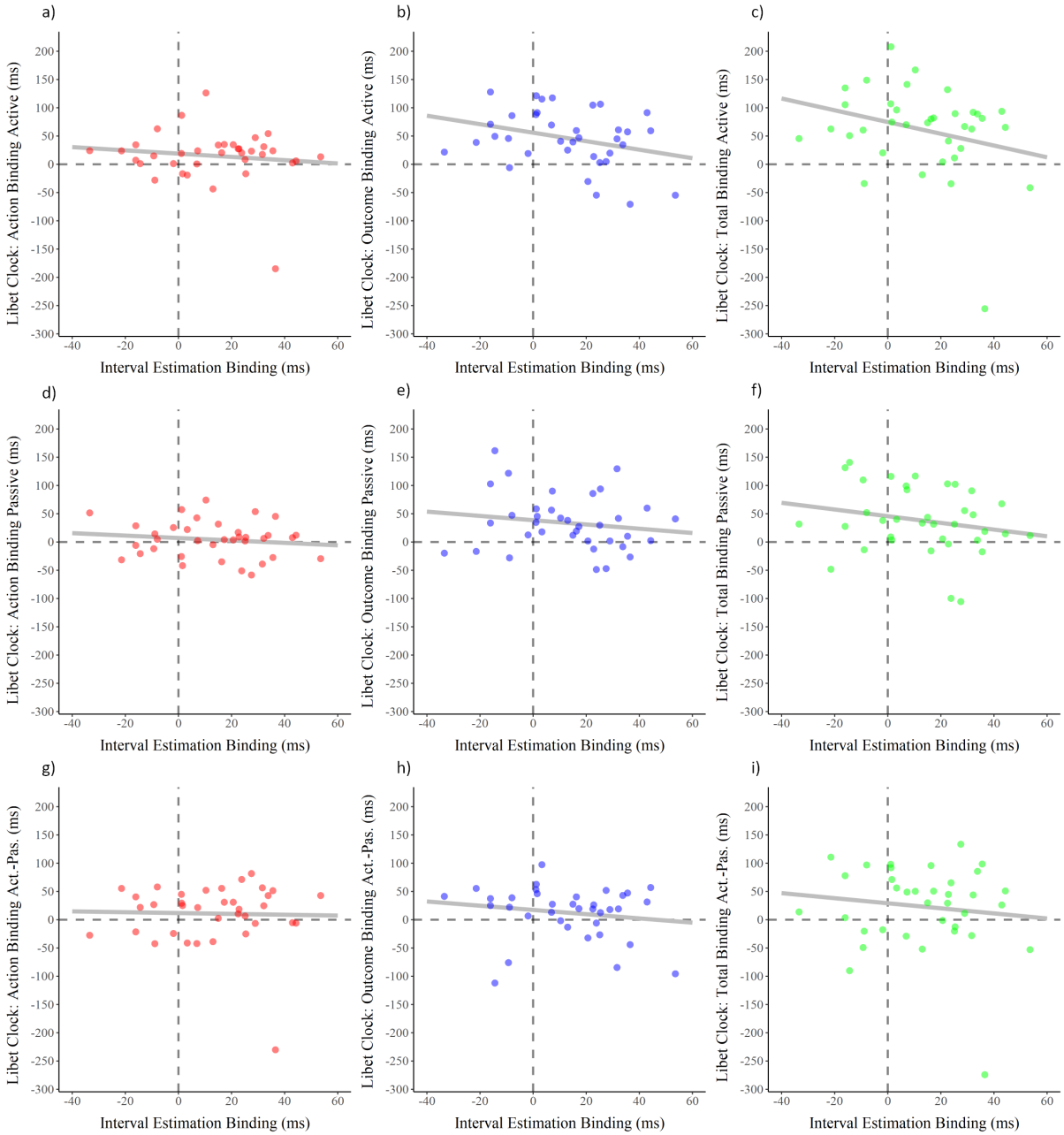
4.2.2 Correlations Between Binding Measures

The last row of Table 2 shows correlations of binding measures between paradigms (visualized in Figure 4). Bayes factors indicate moderate evidence for the null hypothesis, that the true correlation was zero, over the alternative hypothesis, that the true correlation lay above zero. This applies to the correlations between binding in interval estimation and binding in the active conditions (action binding: $t(34) = -0.76$, $p = .774$, $B_{HN(0, 0.454)} = 0.22$, $RR_{B<1/3}$ [0.288, >1.32], outcome binding: $t(34) = -1.89$, $p = .966$, $B_{HN(0, 0.454)} = 0.13$, $RR_{B<1/3}$ [0.166, >1.32], total binding: $t(34) = -1.67$, $p = .948$, $B_{HN(0, 0.454)} = 0.15$, $RR_{B<1/3}$ [0.183, >1.32]), in the passive conditions (action binding: $t(34) = -0.81$, $p = .789$, $B_{HN(0, 0.454)} = 0.22$, $RR_{B<1/3}$ [0.281, >1.32], outcome binding: $t(34) = -0.93$, $p = .821$, $B_{HN(0, 0.454)} = 0.20$, $RR_{B<1/3}$ [0.261, >1.32], total binding: $t(34) = -1.25$, $p = .889$, $B_{HN(0, 0.454)} = 0.17$, $RR_{B<1/3}$ [0.224, >1.32]) and the correlations with the binding values calculated as the difference between active and passive conditions (action binding: $t(34) = -0.17$, $p = .565$, $B_{HN(0, 0.454)} = 0.32$, $RR_{B<1/3}$ [0.434, >1.32], outcome binding: $t(34) = -0.97$, $p = .832$, $B_{HN(0, 0.454)} = 0.20$, $RR_{B<1/3}$ [0.255, >1.32], total binding: $t(34) = -0.73$, $p = .765$, $B_{HN(0, 0.454)} = 0.23$, $RR_{B<1/3}$ [0.295, >1.32]). RRs suggest that these results are somewhat robust except for the correlation between binding in interval estimation and the active-passive difference in action binding. For this correlation, evidence would become inconclusive with a scaling factor below 0.434, corresponding to a prior assumption for a maximum plausible correlation of $r = .70$. The remaining evidence does not turn inconclusive for scaling factors at least above 0.295, corresponding to a maximum plausible correlation of $r = .53$.

There is one interesting result of the exploratory analyses. Regarding binding measures within paradigms, there is a high positive correlation between outcome binding in the active and the passive conditions ($r = .56$) supported by extreme evidence indicated by the associated Bayes factor ($t(34) = -3.92$, $p < .001$, $B_{HN(0, 0.454)} = 213.66$, $RR_{B<1/3}$ [0.060, >1.32]). This result is robust against changes in the prior with only scaling factors below 0.060 leading to other results, which would correspond to a plausible maximum correlation of $r = .12$.

Figure 4

Scatter plots showing the relationship of binding between paradigms in Experiment 2. Displayed pairwise are action, outcome and total binding in the Libet clock paradigm with binding in interval estimation: For the active condition (a-c), the passive condition (d-f) and binding calculated as the difference of the two (g-i). Simple regression lines are provided.



4.3 Preliminary Discussion

4.3.1 Relationship Between Interval Estimation and the Libet Clock Paradigm

The aim of the study was to investigate the relation between the most often used paradigms to measure temporal binding, Libet clock and interval estimation. We hypothesized the two measures to be positively correlated due to broad empirical evidence for time-stable intrapersonal influences on temporal binding and their widespread interchangeable use to measure temporal binding. Our data refutes this hypothesis. Bayes factors revealed moderate evidence against a positive correlation compared to no correlation between paradigms. This holds across different action-outcome delays and for action and outcome binding separately and combined. Study 2 replicates the findings of Study 1 for active conditions in the Libet clock paradigm and expands on them for passive conditions and a binding index analogous to the one in interval estimation, i.e., the difference between the active and passive conditions. These results suggest that the subjective temporal contraction of action and outcome represents different underlying mechanisms in the two paradigms, at least regarding the influence of intrapersonal features. Divergence of underlying mechanisms of different temporal judgements is also proposed by Eagleman (2008).

4.3.2 Reliability and Nomological Implications

The psychometric properties have previously been explored surprisingly little for two widely used methods that proposedly measure temporal binding in a similar fashion. Reliabilities for these measures are usually not reported even if the effect of reduced trial numbers is explored (Cubillas et al., 2020). This, however, would allow a more differentiated interpretation of results of studies that, like ours, found little to no difference in binding between active and passive conditions (e.g., Braun et al., 2014; Poonian et al., 2015; Poonian & Cunnington, 2013; Suzuki et al., 2019). The reliability of the interval estimation task falls into an acceptable range (.72 -.78) in Study 1 and just below that (.63) in Study 2 and changes the interpretation of the result that the average estimated action-onset duration for observed and self-action are equal on a group-level (temporal binding is close to zero). For

example, to obtain a reliability measure in Study 1, we calculated temporal binding, i.e., the difference between active and passive conditions, separately for odd and even trials, creating two halves of trials. Assuming that there is no systematic difference between the conditions in an individual participant, the difference calculated from each half of trials would vary randomly around zero and would not correlate between halves across participants, so the reliability would be very low. Since the difference (the temporal binding measure) is measured somewhat reliably, there seems to be systematic variation due to interindividual differences. Some participants systematically estimated intervals in the self-action condition to be shorter and some participants did the opposite resulting in the group mean to be close to zero. Future research should investigate what these interindividual differences might be. A possible cause could be differences in cue-weighting when processing agency or causality. While one person's processing might be influenced more strongly top-down by knowledge about the causal relationship between action and outcome, another person might rely more on forward motor-processing or proprioceptive action cues. Individual reliance on certain cues has been suggested for schizophrenia (Martin, 2013; Voss et al., 2010).

The reliability for the binding measures in the Libet clock paradigm is higher, suggesting that the method of interval estimation introduces substantial measurement error. A possible solution could be more training trials. Although the number of training trials in our study was comparable to other studies (e.g., Chiarella et al., 2020; Suzuki et al., 2019), there are studies which did not use any training trials (Fereday & Buehner, 2017; Nataraj et al., 2020). Although reliability dropped in Study 2 despite more training in interval estimation, further research could investigate whether using even more training trials, requiring participants to reach a certain performance threshold before starting the experiment or using an analogue scale improves reliability. Aside from trying to improve reliability of binding measures, future research should at least report reliability estimates. Parsons et al. (2019) explain the effects of reliability on statistical power to detect group differences and correlations. The reliability of two measures limits the maximum observable correlation between those measures. This

has direct implications for studies like ours investigating correlations between measures. Additionally, reported reliability can aid interpretation of results across studies and meta-analytical procedures, as Parsons et al. discuss. Another area not covered in our study and in current research for both paradigms is test-retest reliability. The split-half approach employed in our study gives an estimate for a measures internal consistency at a single point in time whereas test-retest reliabilities provide information on a measures stability across multiple assessments (Parsons et al., 2019). It has been implied that temporal binding varies over the life span (Cavazzana et al., 2014, 2017). However, the short-term variation of temporal binding and temporal stability of its measures are to our knowledge not yet researched.

4.3.3 Discussion of Exploratory Analyses

Additional results which are not part of our hypothesis will be discussed purely exploratively. Within-method correlations suggest with moderate evidence, that action and outcome binding in the Libet clock paradigm were uncorrelated in Study 1. This fits partially different proposed mechanisms, particularly sensory pre-activation of action consequences leading to outcome binding (Waszak et al., 2012). Optimal cue integration theory (Synofzik et al., 2013) on the other hand proposes higher action binding when the outcome is a relatively more reliable cue than the action and vice versa. If relative cue reliability weighting would vary across participants, e.g. through differences in access to intentions (Lush et al., 2016, 2019), action and outcome binding should be negatively correlated. Our findings are consistent with those of Tonn et al. (2021) who also did not find a correlation between action and outcome binding in the Libet clock paradigm. If binding in paradigms that do not differentiate action and outcome binding, like interval estimation, represents the sum of the two, one could argue that the resulting score reflects a mixture of underlying mechanisms. This complicates interpretation of what is being measured. However, evidence was inconclusive in Study 2. For another interesting result, we found a large positive correlation between outcome bindings in the active and passive conditions. This suggests that outcome binding results from mechanisms that are part of action

perception regardless of being the agent or observing an agent. This did not seem to be the case for action binding.

4.3.4 Limitations

In Study 1 we employed two temporal binding measures, the Libet clock and interval estimation paradigms, as they have been used in previous research (e.g., De Pirro et al., 2020; Poonian et al., 2015; Poonian & Cunnington, 2013; Schwarz et al., 2019). Additionally, we tried to eliminate differences in task features such as interval length in Study 2. With both approaches we did not find a positive correlation between these binding measures. However, our findings apply only to the circumstances of our study. For example, it would be premature to generalize these findings to action-outcome delays that we did not investigate. Another key feature of our study was the operationalization of active and passive conditions as the participants' and the experimenter's actions respectively. Whether our findings apply to other situations where, for example, participants make voluntary and involuntary movements remains to be seen. We would also caution against simply transferring our findings to other methods of measuring binding such as interval reproduction.

Besides paradigm implementation, generalizability of our results is restricted more generally. Naturally, our results must be replicated especially since Bayes factors indicated only moderate evidence. Further, our sample consisted of healthy young adults studying sports science and as described before, age and disorders can alter temporal binding. How reliable temporal binding measures are in different populations must be investigated in those populations. More generally, reliability of cognitive-behavioral tasks always depends on the specific circumstances under which they are used to measure psychological attributes, including participants, measurement situation and implementation (Parsons et al., 2019). Future research must not rely on our findings but assess their measures' reliability independently.

4.3.5 Conclusion and Future Research

We examined how temporal binding measured by interval estimation and temporal binding in the Libet clock paradigm relate. We found the two measures to be unrelated. From the perspective of nomological networks this raises the question of what exactly is measured by each method and what intrapersonal factors influence each measurement. This connects to recent research which questions whether the proposed measures of so-called temporal binding actually depend on intention (Hoerl et al., 2020; Kirsch et al., 2019; Suzuki et al., 2019). Our results carry two implications for further research: First, our results should be seen preliminarily due to their limitations. Therefore, the measurement of phenomena related to the sense of agency should be further explored psychometrically within and between various aspects of levels that may or may not evade conscious awareness, investigating nomological relations, i.e., construct validity, and reliability. Research on the sense of agency, basic and application-oriented, needs validated and reliable methods to measure it. We believe that psychometric rigor can facilitate research in this field and strengthen confidence in its results. If the mechanisms underlying the various methods of measurement and their relationship are poorly understood, studies that use the different methods must be interpreted cautiously especially when interpreted jointly. Second, the shortcomings of this correlational approach should be complemented experimentally by causal research. If methods that presumably measure the same turn out to be unrelated, understanding of the intrapersonal aspects that influence each respective method might be essential.

5 Study 3: Temporal Binding and Spatial Working Memory: A Correlational Approach[4]

Study 3 aimed to investigate whether binding effects in the Libet clock task could result from processes in spatial working memory. To this end, we tested the following hypotheses. First, we expect to replicate the typical binding effects in the Libet clock task.

H1: Action and outcome binding occur in the Libet clock task.

If binding effects result from processes in spatial working memory, they should also occur when the remembered positions associated with the action and the outcome are not sequenced on a linear, perfectly time-dependent dimension, such as the path of the rotating clock hand.

H2: Action and outcome binding occur in a spatially random task.

If we find binding effects in a spatially random task, we need to investigate whether these effects have the same basis as those in the Libet clock task. If they have, they should correlate across participants. Because action and outcome binding are not correlated, we test this separately for the two binding effects.

H3: Action binding in the Libet clock task correlates positively with action binding in the new, spatially random task.

H4: Outcome binding in the Libet clock task correlates positively with outcome binding in a spatially random task.

If interference between the clock hand positions in spatial working memory leads to the binding effects, having better spatial working memory, especially being better at maintaining and updating,

[4] The results presented in this section were published in advance as Experiment 1 in: Siebertz, M., & Jansen, P. (2025). Is the temporal binding effect in the Libet clock-task based in spatial working memory? A correlational and a dual-task approach. *Consciousness and Cognition*, 134, 103909. https://doi.org/10.1016/j.concog.2025.103909

should reduce binding. The clock hand positions during action and outcome could be kept from interfering with each other more effectively.

H5: Binding in both tasks correlates negatively with spatial working memory performance.

To eliminate more general attentional processes as the cause of binding effects, we expect a correlation with spatial working memory but not with the working memory of other modalities such as visual working memory.

H6: Binding in both tasks does not correlate with visual working memory performance.

The preregistration for Study 3 can be found on OSF (https://osf.io/6prgd). Data, R analysis code and PsychoPy experiment code can be found in the associated repository (doi.org/10.17605/OSF.IO/2NM34).

5.1 Method

5.1.1 Design

Study 3 uses a correlational design to investigate the theory that temporal binding in the Libet clock task originates from spatial memory processes. All participants complete four tasks: The Libet clock task, a 2-back task with spatial stimuli, a 2-back task with visual stimuli, and a task resembling the Libet clock task. The latter replaces the clock face with a circle in which dots appear simultaneously with the action and outcome and whose positions must later be recalled.

5.1.2 Sample

To achieve a sufficiently large sample size, we used adaptive sampling with hypothesis-relevant Bayes factors below 1/3 or above 3 as a stopping criterion. To set a viable endpoint for sampling if Bayes factors fail to indicate conclusive evidence, we conducted a power analysis in G*Power (version 3.1.9.7, Faul et al., 2009). For our correlational hypotheses, to achieve a power of at least 95% to detect differences from zero for correlations of .3 or -.3 at a significance level of 5% in a one-sided test requires 111 participants. We decided that the effect size of .3 would be the smallest effect of interest.

To account for exclusions based on our previous experience, we increased this sample size by 10% resulting in 122 participants. As preregistered, we first inspected the data at 50 participants and then after every 10 new participants.

Because some Bayes factors failed to indicate conclusive evidence at all steps, we sampled all 122 participants. Of those, seven reported current neurological or psychiatric disorders. One reported impaired and uncorrected vision and hearing respectively. Due to these preregistered criteria, nine participants were excluded from further analyses. Following the preregistered workflow, nine other participants were excluded for having judgment errors 3 *SD* above or below the condition sample mean in at least one condition of the two binding tasks. This left a final sample of 104 participants, 50 reported their gender as female, 54 as male and none as non-binary. Mean age was 22.9 years (*SD* = 3.4). All participants were healthy, young students of sports science at our faculty, gave written informed consent and received course credit for participation.

The study was executed in accordance with the declaration of Helsinki for the guidelines of ethical considerations. Ethical approval for this study was not required in accordance with the conditions outlined by the German Research Society (DFG) where research that carries no additional risk beyond daily activities does not require a research ethics board's approval. We communicated all considerations necessary to assess the question of ethical legitimacy of the study.

5.1.3 Material

Participants worked through the tasks and the demographic questionnaire in PsychoPy (version 2023.1.1, Peirce et al., 2019) on a Lenovo Thinkpad T15 wearing noise-cancelling headphones. Responses were given using an external keyboard and a mouse with a mouse wheel. The demographic questionnaire queried age, gender, and the exclusion criteria of current psychiatric or neurological disorders and uncorrected, impaired vision and hearing.

5.1.3.1 Libet Clock Task

The Libet clock task is identical to the one in Study 1 which we adapted from Kirsch et al. (2019), and goes back to the study of Haggard et al. (2002). Participants observed a clock face with a clock hand rotating at 2560 ms per revolution after starting at a random position. In the contingent conditions the participants' action, a key press, was followed by an outcome, a sine tone (1000 Hz, 100 ms duration) with a delay of 250 ms. Figure 5 shows the sequence of events for a contingent trial. Participants' task was to press the space bar at a time of their chosing after the first revolution. This triggered the tone. Beforehand, they were instructed to remember the clock hand position during their key press or during the tone. The clock hand continued to rotate after the target event for 1500 to 2500 ms before the clock disappeared for 1000 ms. The clock hand reappeared randomly positioned 45° to 60° before or after its position during the target event. Participants used the mouse wheel to turn the clock hand to the position they remembered.

In the two baseline conditions, participants either performed a key press that did not trigger a tone or remained passive, and the tone was played automatically 2810 ms to 4810 ms after the start of the trial. After the clock disappeared for 1000 ms, participants reinstated the clock hand position during the target event. The baseline condition with the tone as the target event mirrored its contingent counterpart with 2560 ms for the first revolution, 250 ms action-outcome delay and 2000 ms jitter. The order of the four conditions was counterbalanced across participants. There were 28 trials per condition, and participants did three unanalyzed training trials per condition in the same condition order before the actual task started.

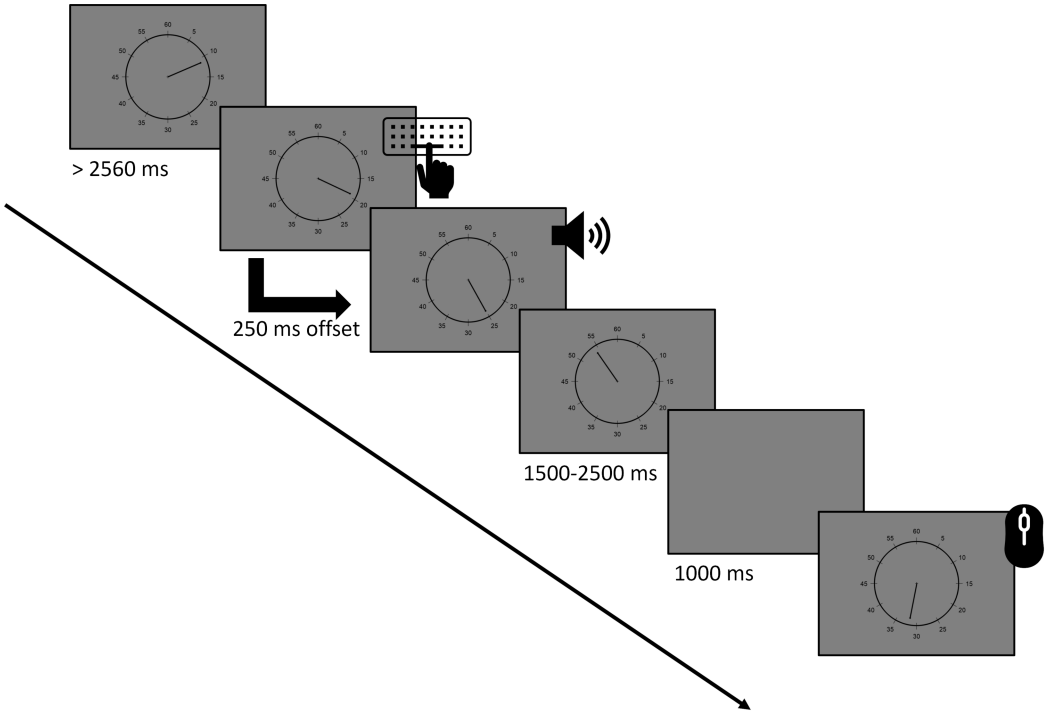
We calculated temporal binding scores for the action as the mean judgment errors of the contingent minus the baseline condition. The binding score for the outcome is calculated the opposite way which leads to higher, more positive scores to indicate more binding for action and outcome. For comparability to the spatially random binding task described below, we report binding magnitude in proportion to the actual spatial distance between the clock hand positions during the two events

instead of the usual time shift metric. With the 250 ms action-outcome delay corresponding to 35.16° of rotation, this equals the proportion of the actual temporal distance.

We calculated reliability scores for action and outcome binding by calculating the split-half reliability. We used 5000 random permutations of splitting the trials per condition into two halves and calculated the action and outcome binding scores each time (Parsons et al., 2019). Averaged across the permutations, Spearman-Brown attenuation-corrrected split-half reliability as the correlation between binding scores for the two halves was $r_{split\text{-}half}$ = .86 (95% CI = [.81, .90]) for action binding and $r_{split\text{-}half}$ = .98 (95% CI = [.98, .99]) for outcome binding.

Figure 5

The sequence of events in a contingent trial of the Libet clock task. Depending on the target event, participants reinstate the clock hand position during the key press or the tone using the mouse wheel at the end of the trial.



5.1.3.2 Spatially Random Binding Task

The Libet clock task confounds the temporal and spatial distance between action and outcome as they correlate perfectly. Through the constant rotation speed, a certain temporal delay between action

and outcome corresponds to a fixed angle of rotation in the clockwise direction. To disentangle the temporal and the spatial delay, we designed a task that mimics the Libet clock task. Instead of positions of a clock hand, participants observe pairs of consecutively appearing dots in a circle.

The task contained the same four conditions as the Libet clock task, two contingent and two baseline conditions in which participants had to remember the position of a dot appearing simultaneously with their key press or a thereby triggered tone. Figure 6 shows an exemplary trial. Participants saw a circle in which a randomly changing greyscale visual noise mask flickered at 60 Hz. The mask's purpose was to introduce visual noise comparable to the motion of the clock hand in the Libet clock task. In the contingent conditions, participants pressed the space bar at a time of their choice, immediately letting a dot appear at a random position inside the circle for 100 ms. After a delay of another 150 ms after the disappearance of the dot a sine tone (1000 Hz, 100 ms duration) was played leading to 250 ms of action-outcome delay like in the Libet clock task. Simultaneously with the tone a second dot appeared for 100 ms. The distance between the two dots was equal to the distance between the position of the tip of the clock hand during the key press and the tone in the Libet clock task. The direction in which the second dot appeared relative to the first one was random. Immediately after the disappearance of the second dot, the visual noise mask was displayed for 500 ms to prevent afterimages from highlighting the dot positions and to introduce visual noise. Then, the mouse cursor appeared, and participants clicked where they recalled the position of the dot appearing with the current target event.

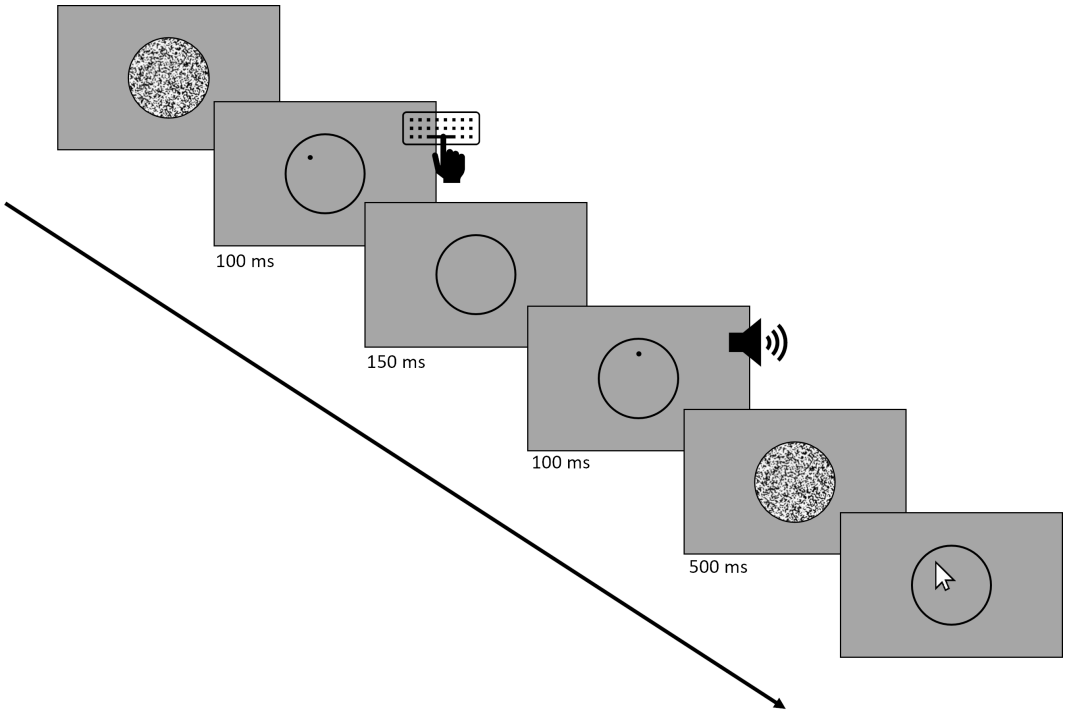
In the baseline condition with the key press as the target event, no second dot appeared, and no tone was played. Participants did not press the space bar in the baseline condition with the target event tone. The first dot appeared automatically after a random delay of 750 ms to 1750 ms after trial onset. There were 28 trials and three unanalyzed training trials per condition. Participants did the latter before the actual task in the same condition order. Condition order was counterbalanced across participants.

The judgment error was calculated by orthogonally projecting the position that participants clicked at onto the line going through both actual dot positions. The judgment error was the distance of this projection to the actual target dot position. The 28 pairs of dot positions were generated randomly once at the start of the task and then reused in random order for all conditions to control for the positioning of the dots relative to the bordering circle. Judgment errors in the baseline conditions were calculated as described based on the virtual position of the non-presented dot relative to the presented dot. As with the Libet clock task, the difference between the judgment errors in the baseline and contingent conditions for each target event yielded the action and outcome binding scores. Higher, more positive values would indicate higher binding for both, which would mean that participants remembered the dot position systematically shifted towards the other dot in the contingent compared to the baseline condition.

Using the permutation-based split-half approach for calculating reliability described for the Libet clock task, the average spearman-brown attenuation corrected split-half reliability was $r_{\text{split-half}} = .65$ (95% CI = [.54, .74]) for action binding and $r_{\text{split-half}} = .74$ (95% CI = [.66, .81]) for outcome binding.

Figure 6

The sequence of events in a contingent trial of the spatially random task.



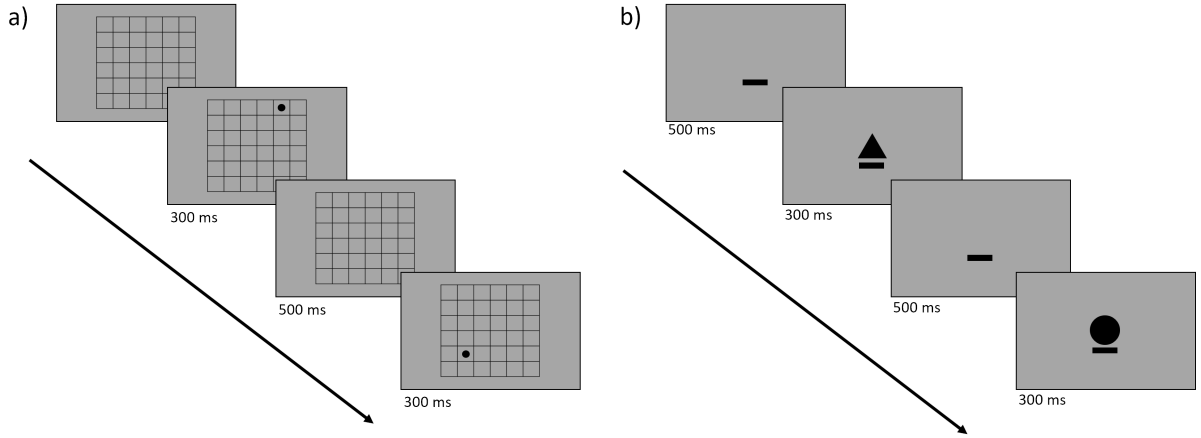
5.1.3.3 Spatial 2-Back Task

To measure spatial working memory performance, we used a 2-back task in which participants had to remember the positions of dots appearing sequentially in a 6-by-6 grid (see Figure 7a). The dots stayed visible for 300 ms with an ISI of 500 ms. Participants were instructed to press the space bar when a dot appeared in the same field as the next-to-last dot. Per block participants saw a sequence of 25 dots, five of which were 2-back repetitions. Participants started with two training blocks with a stimulus duration of 500 ms. If participants reacted in these blocks, the sequence stopped, and feedback was displayed until the participant continued the sequence via key press. Two further training blocks followed, which were identical to the following six actual task blocks with a stimulus duration of 300 ms and no feedback about the correctness of their response given. Training blocks were not analyzed. Participants could react to a stimulus while it was visible and during the ISI until the next stimulus was presented. To provide them with feedback that their key press was registered in time, the outer border of the grid changed color to a lighter grey for 100 ms. Sequences were equal

and presented in the same order for all participants to eliminate sequence order as a source of variance.

Figure 7

The sequence of events of the spatial (a) and visual (b) 2-back tasks.



For our analyses, we calculated *d*' as an informative index for working memory performance (Haatveit et al., 2010). Across the six blocks, participants saw 150 stimuli of which 30 were targets. We calculated the *z*-transformed hit rate and false alarm rate for each participant and subtracted the latter from the former to derive *d*'. Hit rates of 1 and false alarm rates of 0 were replaced with $1 - 1 / (2 * n_{targets})$ and $1 / (2 * n_{non-targets})$ respectively. Again, we used the permutation-based split-half approach to calculate reliability scores for the *d*'-index. Splitting targets and non-targets into halves separately using 5000 permutations and calculating *d*' for each half, the average spearman-brown corrected inter-half correlation was $r_{split\text{-}half} = .67$ (95% CI = [.58, .75]).

5.1.3.4 Visual 2-Back Task

The 2-back task measuring visual working memory was identical to the one for spatial working memory apart from the stimulus material. Instead of dots in a grid, participants had to remember geometric shapes presented at the center of the screen (see Figure 7b). There were 5 shapes: A circle, a square, a triangle, a star and a pentagon. To provide feedback for key press registration, a black bar

was displayed below the position where the shapes appeared. Like the grid border in the spatial 2-back task, it changed color to a lighter grey for 100 ms upon key press.

We calculated the *d*'-index as described for the spatial working memory task. The same permutation-based split-half approach resulted in an average spearman-brown corrected inter-half correlation of $r_{split\text{-}half}$ = .57 (95% CI = [.45, .69]) for *d*'.

5.1.4 Procedure

First, participants read the study information and gave written informed consent. They started the testing session with the spatial and visual 2-back tasks. The task order was counterbalanced across participants. After that, participants did the two binding tasks. Again, task order was counterbalanced across participants. Finally, they answered the demographic questionnaire. Participants were encouraged to ask questions at any time and proceed through the testing session at their own pace. Testing sessions typically lasted for one to one and a half hours.

5.1.5 Statistical Analysis

We used Bayes factors for population-level inference to investigate the absence of an effect in Hypothesis 6 and to use adaptive sampling. For Hypotheses 1 and 2, expecting binding above zero in both binding tasks, we calculated Bayes factors by comparing the null model (H0) with the alternative (H1). H0 was modeled as a normal distribution with a mean of zero and a *SD* equal to the observed *SD* of the binding in question. H1 was modeled as a half-normal distribution with mode 0 and *SD* equal to the expected effect. Using data from Study 1, this was 13 ms for action binding and 74 ms for outcome binding. These values correspond to 5.2% and 29.6% of the actual delay of 250 ms. Therefore, in proportion of the actual event position distance, the *SD* of the H1 model was set to 0.052 and 0.296 respectively.

For Hypotheses 3 and 4, we calculated Bayes factors using the Fisher's *z*-transformation of the Pearson correlation coefficient. The H0-model was a normal distribution with mean zero and $SD = 1/\sqrt{df - 1}$. The alternative H1 was modeled as a half-normal distribution with mode 0 and *SD*

equal to the expected correlation effect size. Following Dienes (2019), this was set to half of a rough sensible maximum. Using reliability estimates from Study 1 and accounting for reliability-related correlation attenuation. we calculated the maximum observable correlation of binding between the two tasks using the formula: $r_{expected} = r_{theoretical} * \sqrt{rel(x) * rel(y)}$, with $r_{expected}$ being the observable correlation attenuated by imperfect reliability, $r_{theoretical}$ being the true correlation between the variables and *rel*(x) and *rel*(y) being the reliabilities of the variables. Assuming that binding in the two tasks is caused by the same mechanisms, the $r_{theoretical}$ was set to 1. Reliability was assumed to be .84 for action binding and .96 for outcome binding. Using the formula above, these were simultaneously the maximum expected correlations. Accordingly, the *SD* of the half-normal H1 model for the between-task correlation of action binding was set to $Fisher's\ z(.84)/2 = 0.612$ and the *SD* for outcome binding to $Fisher's\ z(.96)/2 = 0.973$.

Hypotheses 5 and 6 regard the correlation between working memory performance and binding. The models for H0 and H1 were constructed as described for Hypotheses 3 and 4. To investigate even small correlations, we compared the null model that assumes the correlation is zero with an alternative that assumes the correlation is -.3. As it is computationally equivalent, the observed correlation's sign was flipped and the H1 half-normal was modeled with an *SD* of $Fisher's\ z(.30) = 0.310$.

We calculated robustness regions to check the dependence of our results on our prior specifications. For mean comparisons, Bayes factors were calculated for prior effect sizes from 0.001 to five times the actual specified effect size. The interval in which the same qualitative result regarding the Bayes factor (<1/3, >3 or between) held true is reported as the robustness region. Bayes factors were calculated for *SD*s corresponding to *r* = .001 ($SD = Fisher's\ z(r) = 0.001$) to *r* = .999 ($SD = Fisher's\ z(r) = 3.800$) for correlations.

5.2 Results

5.2.1 Action and Outcome Binding in the Libet Clock Task

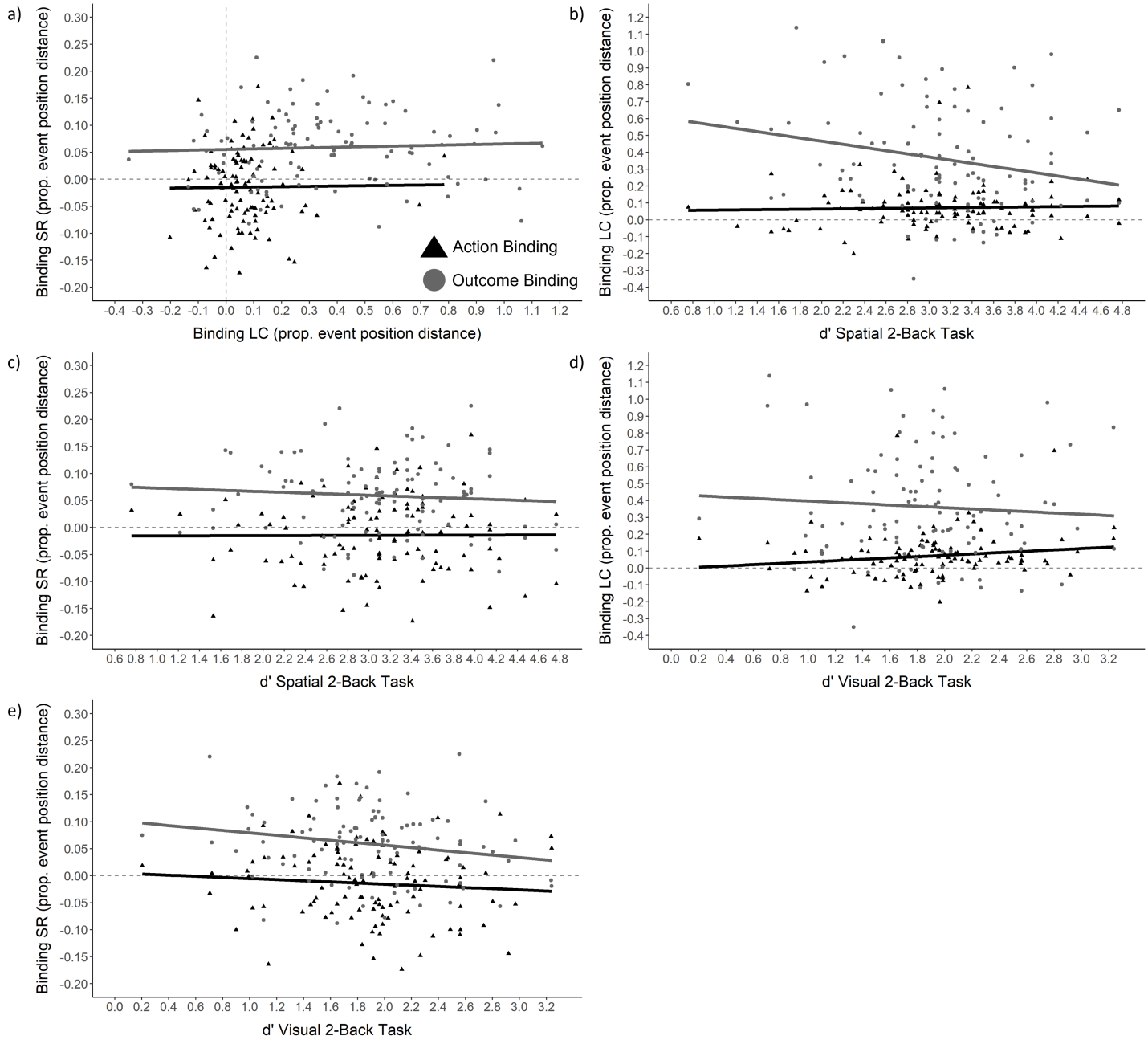
Figure 8a shows all binding scores. Average action binding in the Libet clock task was 7.1% of the event position distance (*SD* = 13.4%, corresponding to a temporal shift of 17.8 ms, *SD* = 33.5 ms). The Bayes factor shows overwhelming evidence for action binding being above zero ($t(103) = 5.44$, $p < .001$, $B_{HN(0,\ 0.052)} >> 100$, $RR_{B>3}$ [0.003, >0.26]). Mean outcome binding was 36.3% of event position distance (*SD* = 31.0%, 90.8 ms and *SD* = 33.5 ms in temporal shift). Again, the Bayes factor shows overwhelming evidence for outcome binding above zero ($t(103) = 11.92$, $p < .001$, $B_{HN(0,\ 0.296)} >> 100$, $RR_{B>3}$ [0.003, >1.48]). These results replicate the typical binding effect in the Libet clock task and confirm Hypothesis 1. Robustness regions indicate virtually no dependence on prior effect size. The correlation between action and outcome binding in the Libet clock task was $r = .03$ (95% CI = [-.17, .22]).

5.2.2 Action and Outcome Binding in the Spatially Random Task

Mean action binding in the spatially random task was -1.4% of event position distance (*SD* = 7.9%). The Bayes factor shows strong evidence for action binding not being above zero ($t(103) = -2.14$, $p = .983$, $B_{HN(0,\ 0.052)} = 0.04$, $RR_{B<1/3}$ [0.006, >0.26]). On average, outcome binding was 5.9% of event position distance (*SD* = 6.6%). The Bayes factor indicates overwhelming evidence for outcome binding above zero ($t(103) = 9.08$, $p < .001$, $B_{HN(0,\ 0.296)} >> 100$, $RR_{B<3}$ [0.003, >1.48]). This confirms the Hypothesis that binding effects occur in a spatially random task only for outcome binding and rejects it for action binding. Large robustness regions indicate virtually no dependence on prior effect size. The correlation between action and outcome binding in the Libet clock task was $r = .07$ (95% CI = [-.12, .26]).

Figure 8

Scatter plots showing the relationships between binding in the Libet clock task and the spatially random binding task (a), spatial working memory performance and binding in the Libet clock (LC) task (b) and the spatially random (SR) binding task (c), and visual working memory performance and binding in the Libet clock task (d) and the spatially random binding task (e). Data regarding action binding is shown in black triangles, and regarding outcome binding is in grey circles. Regression lines show the results of simple linear regression.



5.2.3 Relationship of Binding Between Binding Tasks

Figure 8a shows the relationship of binding scores between the two binding tasks. The correlation between action binding scores was .01 (95% CI = [-.18, .20]). The Bayes factor indicates moderate evidence for the absence of a positive correlation ($t(102) = 0.12$, $p = .451$, $B_{HN(0,\ 0.611)} = 0.18$, $RR_{B<1/3}$ [0.316, 1.323]). These robustness region's limits correspond to assumed rough maximum correlations of .56 to .99. This result rejects Hypothesis 3. Compared to the assumed rough maximum correlation of .84, the robustness region indicates little dependence on prior effect size on the lower end.

The correlation between outcome binding scores was .05 (95% CI = [-.15, .24]). Again, the Bayes factor indicates moderate evidence for the absence of a positive correlation ($t(102) = 0.48$, $p = .314$, $B_{HN(0,\ 0.973)} = 0.16$, $RR_{B<1/3}$ [0.444, 1.323]). The robustness region's limits correspond to rough maximum correlations of .71 to .99. Hypothesis 4 is rejected. Compared to the assumed rough maximum effect size of .96, the robustness region indicates some dependence on the prior effect size on the lower end.

5.2.4 Relationship Between Binding and Working Memory Performance

The average *d*' in the spatial 2-back task was 3.1 (*SD* = 0.7, range = [0.8, 4.8]). Average *d*' in the visual 2-back task was 1.9 (*SD* = 0.5, range = [0.2, 3.2]). The correlation between spatial and visual working memory performance was $r = .31$ (95% CI = [.13, .48]). Table 3 shows the results of our analyses regarding the correlation between binding and spatial and visual working memory performance.

Table 3

Correlations between binding scores in the Libet clock and spatially random tasks with spatial and visual working memory performance in the 2-back task. Bayes factors over 1 favor a negative correlation versus no correlation.

			95% CI					RR		
	Binding type	*r*	LL	UL	*t*(102)	$p_{r<0}$	$B_{HN(0,\ 0.310)}$	Qual. res.	LL	UL
SWM	Action LC	.04	-.23	.16	0.36	.641	0.24	*B*<1/3	0.107	1.323
	Outcome LC	-.23	-.40	-.03	-2.34	.011	6.76	*B*>1/3	0.030	0.454
	Action SR	.01	-.19	.20	0.05	.520	0.29	*B*<1/3	0.138	1.323
	Outcome SR	-.07	-.26	.12	-0.75	.228	0.60	1/3<*B*<3	0.000	0.295
VWM	Action LC	.16	-.03	.34	1.65	.949	0.12	*B*<1/3	0.055	1.323
	Outcome LC	-.07	-.26	.13	-0.70	.244	0.57	1/3<*B*<3	0.000	0.275
	Action SR	.08	-.11	.27	-0.83	.203	0.66	1/3<*B*<3	0.000	0.324
	Outcome SR	-.19	-.37	.01	-1.92	.029	3.01	*B*>1/3	0.050	0.155

Note. RR: Robustness region, SWM: Spatial working memory performance, VWM: Visual working memory performance, LC: Libet clock task, SR: spatially random task, Qual. res. = qualitative result.

5.2.4.1 Relationship Between Binding and Spatial Working Memory Performance

The correlation between action binding in the Libet clock task and spatial working memory performance was .04 (95% CI = [-.23, .16], black triangles in Figure 8b). The Bayes factor indicates moderate evidence against the correlation being below zero ($t(102) = 0.36$, $p = .641$, $B_{HN(0,\ 0.310)} = 0.24$, $RR_{B<1/3}$ [0.107, 1.323]). Robustness region limits correspond to rough maximum correlations of -.21 to -.99.

The correlation between outcome binding in the Libet clock task and spatial working memory performance was -.23 (95% CI = [-.40, -.03], grey circles in Figure 8b). The Bayes factor shows moderate evidence in favor of the correlation being below zero ($t(102) = 2.34$, $p = .011$, $B_{HN(0,\ 0.310)} = 6.76$, $RR_{B>3}$ [0.030, 0.454]). Robustness region limits correspond to rough maximum correlations of -.06 to -.72.

The correlation between action binding in the spatially random task and spatial working memory performance was .01 (95% CI = [-.19, .20], black triangles in Figure 8c). The Bayes factor shows moderate evidence against the correlation being below zero ($t(102) = 0.05$, $p = .520$,

$B_{HN(0, 0.310)}$ = 0.29, $RR_{B<1/3}$ [0.138, 1.323]). Robustness region limits correspond to rough maximum correlations of -.27 to -.99.

The correlation between outcome binding in the spatially random task and spatial working memory performance was -.07 (95% CI = [-.26, .12], grey circles in Figure 8c). The Bayes factor shows inconclusive evidence regarding the correlation being below zero (t(102) = -0.75, p = .228, $B_{HN(0, 0.310)}$ = 0.60, $RR_{1/3<B<3}$ [0.000, 0.295]). Robustness region limits correspond to rough maximum correlations of .00 to -.53. Prior effect sizes below -.53 lead to the Bayes factor showing moderate evidence against the correlation being below zero.

5.2.4.2 Relationship Between Binding and Visual Working Memory Performance

Action binding in the Libet clock task correlated with visual working memory performance at r = .16 (95% CI = [-.03, .34], black triangles in Figure 8d). This yields substantial evidence against the correlation being below zero (t(102) = 1.65, p = .949, $B_{HN(0, 0.310)}$ = 0.12, $RR_{B<1/3}$ [0.055, 1.323]). The robustness region indicates the same qualitative results for prior correlation effect sizes between .11 and .99.

Outcome binding in the Libet clock task showed a correlation of r = -.07 (95% CI = [-.26, .13], grey circles in Figure 8d). Evidence for the correlation being below zero is inconclusive (t(102) = -0.70, p = .244, $B_{HN(0, 0.310)}$ = 0.57, $RR_{1/3<B<3}$ [0.000, 0.275]). The robustness region corresponds to prior correlation effect sizes between .00 and .50, with prior effect sizes above .50 leading to conclusive evidence against the correlation being below zero.

Action binding in the spatially random task was correlated with visual working memory performance with an r of .08 (95% CI = [-.11, .27], black triangles in Figure 8e). Evidence for a correlation below zero is inconclusive (t(102) = -0.83, p = .203, $B_{HN(0, 0.310)}$ = 0.66, $RR_{1/3<B<3}$ [0.000, 0.324]). Robustness region limits correspond to prior correlation effect sizes of .00 and .57 with larger prior effect sizes leading to evidence against a correlation below zero.

Outcome binding in the spatially random task correlates with spatial working memory performance at r = -.19 (95% CI = [-.37 .01], grey circles in Figure 8e). This results in moderate evidence in favor of the correlation being below zero ($t(102)$ = -1.92, p = .029, $B_{HN(0,\ 0.310)}$ = 3.01, $RR_{B>1/3}$ [0.050, 0.155]). The robustness region limits translate to correlation prior effect sizes of .10 and .30, with prior effect sizes below and above leading to inconclusive evidence.

5.3 Preliminary Discussion

Study 3 aimed to investigate whether binding in the Libet clock task is rooted in spatial working memory processes. We compared binding in the Libet clock task with binding in a similar task where action- and outcome-related event locations are not in a linear and predictable spatial relationship. We also investigated whether binding in the two tasks is reduced with better spatial working memory performance but not with visual working memory performance.

We were able to replicate the typical binding effects in the Libet clock task confirming Hypothesis 1. Binding magnitudes are in the usual magnitude and ratio with mean action binding of 18 ms and mean outcome binding of 91 ms. Reliability is good for both and better for outcome binding in accordance with previous studies.

Average action binding in the spatially random task is not significantly above zero and does not correlate with action binding in the Libet clock task. Outcome binding in the spatially random task is significantly above zero but only of similar magnitude as action binding in the Libet clock task. However, despite finding a systematic shift in the reported outcome-related position towards the action-related one, it is not correlated to outcome binding in the Libet clock task. This confirms Hypothesis 2 that binding occurs in the spatially random task, but only for outcome binding. Hypotheses 3 and 4 regarding correlations between the two binding tasks are rejected. Reliability is lower than in the Libet clock task at .65 and .74 for action and outcome binding respectively. This is

especially interesting for action binding. Despite the sample average being close to zero, individuals consistently biased their judgments.

Even though we found evidence for binding in the spatially random task, it seems to result from different processes than binding in the Libet clock task. Binding in the Libet clock task may depend on the specific judgment modality of the clock face. Another explanation could be that the visual noise introduced in the spatially random task did not lead to similar perceptual uncertainty compared to the movement of the clock hand. This could especially be the case, as there was no visual noise between action and outcome whereas the clock hand moves between action and outcome in the Libet clock task. This explanation fits the higher variance of action and outcome binding in the Libet clock task compared to the spatially random task.

To measure spatial and visual working memory performance we used 2-back tasks. We expected binding to correlate negatively with spatial working memory performance (Hypothesis 5). This could only be confirmed for outcome binding in the Libet clock task. Hypothesis 5 is rejected for action binding in both tasks, and evidence for outcome binding in the spatially random task is inconclusive. Hypothesis 6 states that binding does not correlate with visual working memory performance. This could be shown for action binding in the Libet clock task. Evidence remains inconclusive for outcome binding in the Libet clock task and action binding in the spatially random task. Outcome binding in the latter does seem to decline with increasing visual working memory performance.

As action and outcome binding in the Libet clock task are uncorrelated and, therefore, stem from different processes, it is not surprising that only one binding type correlates with spatial working memory performance. Our results might be explained by outcome binding in the Libet clock task stemming from interference in spatial working memory. Better spatial working memory abilities would then lead to a more effective shielding of the relevant outcome-related position from the preceding action-related position via updating of the latter. Arguably, the 2-back task consists of a higher

updating to maintaining ratio than for example a 3-back task (Ren et al., 2023). A spatial working memory task relying more on maintaining relative to updating might show a relationship to action binding.

Our results on visual working memory performance do not allow us to completely exclude more general attentional mechanisms as an alternative explanation. As the evidence remains inconclusive, especially for outcome binding in the Libet clock task, we cannot rule out a relationship with visual working memory processes comparable to that with spatial working memory.

6 Study 4: Temporal Binding and Working Memory: A Dual Task Approach[5]

In Study 3 we took a correlational approach. To test the theory that temporal binding results from processes in spatial memory more thoroughly, we expand this with an experimental approach in Study 4. Our theory postulates that binding in the Libet clock task occurs because the clock hand positions during action and outcome interfere with each other as spatial memory content. Accurate recall of the target clock hand position should rely on the use of limited attentional resources to shield it from interference. Based on sensorimotor recruitment models such as attention-based rehearsal (D'Esposito & Postle, 2015) we assume that loading spatial memory binds those resources. This should impair the ability to prevent interference by the irrelevant clock hand position and increase the binding effect. If temporal binding in the Libet clock task originates specifically from spatial memory processes instead of more general attentional processes, loading other memory modalities like visual memory for objects should not increase binding. These considerations lead us to the following hypotheses. Hypothesis 1 is visualized in Figure 9a and Hypothesis 2 in Figure 9b.

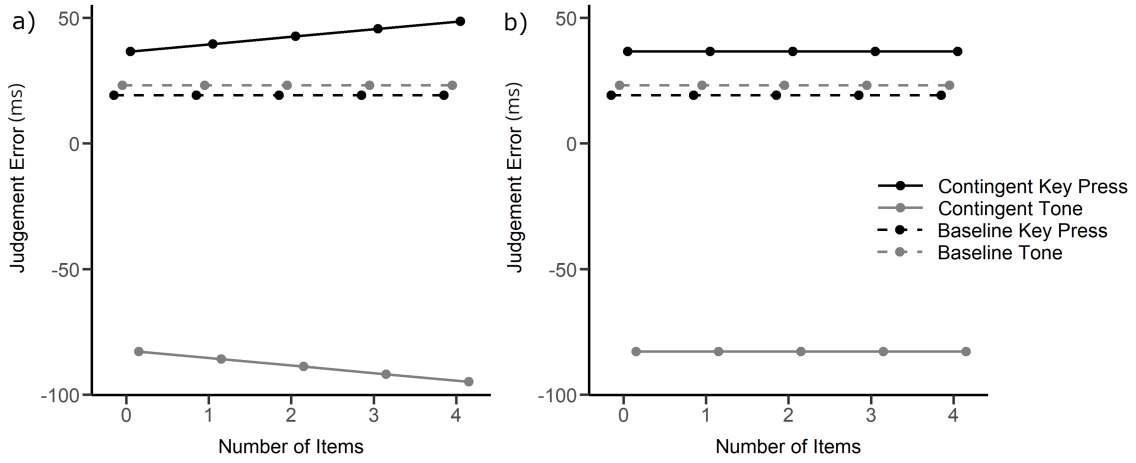
H1: In the Libet clock task embedded in the spatial memory task, temporal binding increases the more items have to be remembered.

H2: In the Libet clock task embedded in the visual object-related memory task, the number of items does not increase temporal binding.

[5] The results presented in this section were published in advance as Experiment 2 in: Siebertz, M., & Jansen, P. (2025). Is the temporal binding effect in the Libet clock-task based in spatial working memory? A correlational and a dual-task approach. *Consciousness and Cognition*, 134, 103909. https://doi.org/10.1016/j.concog.2025.103909

Figure 9

Visualization of the hypotheses for Study 4. Action and outcome binding are expected to increase with spatial memory load (a) but not with visual memory load (b). Binding manifests in the difference between the contingent (solid lines) and the baseline conditions (dashed lines) for the action (target event key press, black lines) and the outcome (target event tone, grey lines).



The preregistration can be found on OSF (https://osf.io/vxagf). Data, R analysis code and PsychoPy experiment code for Study 4 can be found in its OSF repository (doi.org/10.17605/OSF.IO/5FK7Z).

6.1 Method

6.1.1 Design

Study 4 utilized an experimental within-participant design with four factors. Participants did the Libet clock task embedded in the retention interval of a memory task. We manipulated the type of memory stimulus (spatial vs. visual) and the number of items participants had to remember (0-4). Like Study 3 and as usual in the Libet clock task, participants did two baseline and two contingent conditions with the target event being the action and the outcome respectively.

6.1.2 Sample

Our sample size was derived from a simulation-based power analysis (see OSF repository for R code). Data was simulated 5000 times with a *SD* of 1, a slope for condition of 0.3 and a slope for number of items of 0.05 in the contingent condition and 0.0 in the baseline condition. These effect sizes were chosen to resemble a small effect size with smaller effects being not of interest to our hypotheses. For

a linear regression model including condition (contingent vs. baseline), number of items and their interaction, the power analysis resulted in 95 participants to reach a power of .90 for the interaction at a significance level of 5%. Our preregistered analysis using Bayesian mixed models should lead to a higher power. To account for drop-out based on our previous studies of temporal binding, we increased this by 10%, leading to a target sample size of 105 participants.

Six participants reported current neurological or psychiatric disorders, four impaired and uncorrected hearing, and three impaired and uncorrected vision. With overlaps, eleven participants were excluded due to these preregistered criteria leading to a final sample size of 94 participants. 45 participants reported their gender to be female, 49 male and none non-binary. On average, participants were 22.13 years old (*SD* = 2.98). All participants were healthy, young students of sports science at our faculty and received course credit for participation.

The study was executed in accordance with the declaration of Helsinki for the guidelines of ethical considerations. Ethical approval for this study was not required in accordance with the conditions outlined by the German Research Society (DFG) where research that carries no additional risk beyond daily activities does not require a research ethics board's approval. We communicated all considerations necessary to assess the question of ethical legitimacy of the study.

6.1.3 Material

Participants worked through the memory task-embedded Libet clock task and answered the demographic questionnaire in PsychoPy (version 2023.1.3, Peirce et al., 2019) on a Lenovo Thinkpad T15 wearing noise-cancelling headphones over which they heard the auditory stimuli. They responded using an external keyboard and a mouse with mouse wheel. The demographic questionnaire queried age, gender and the exclusion criteria of current psychiatric or neurological disorders and uncorrected, impaired vision and hearing.

The experiment consisted of two instances of the Libet clock task, each embedded in the retention interval of a memory task. The two Libet clock tasks were identical besides the stimulus

material of the embedding memory task. Each trial started with the presentation of the memory stimuli immediately followed by one trial of the Libet clock task. After the presentation and response phases of the Libet clock task participants had to recognize the presented memory stimuli. All PsychoPy code is available in the OSF repository.

6.1.3.1 Libet Clock Task

The Libet clock task was identical to the one in Study 3, but each trial was embedded in the retention interval of one of two memory tasks. Condition order was again counterbalanced across participants and was the same for both instances of the task. Each condition consisted of 30 trials, 6 for each number of items in the memory task (0-4). Before the presentation phase of the memory task, participants saw a screen informing them of the current Libet clock condition and target event, how many trials of the current condition they already completed and how many conditions were left. They could start the trial via key press. Before the two instances of the Libet clock task, participants did three trials per condition without the embedding memory tasks in the same condition order that was used in the actual runs. Training trials were not analyzed.

To examine reliability of temporal binding scores we calculated the correlation between binding scores during the spatial and the visual memory task for action and outcome binding across participants respectively. For action binding this was $r_{\text{test-retest}}$ = .64 (95% CI = [.50, .74]) and for outcome binding $r_{\text{test-retest}}$ = .81 (95% CI = [.72, .87]). Like in our studies reported before, we found a higher reliability for outcome binding.

6.1.3.2 Spatial Memory Task

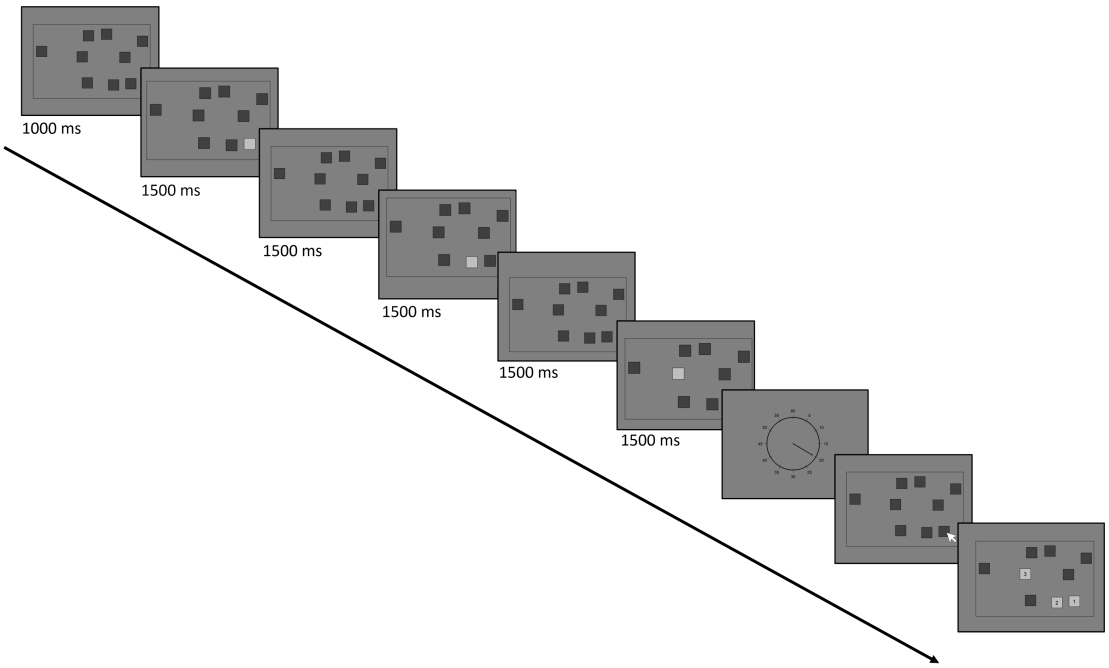
The spatial memory task was adapted from the Corsi block-tapping test (Corsi, 1972). Figure 10 shows an exemplary trial. At the start of each trial, participants saw nine dark grey squares on a grey background randomly positioned inside a rectangular border for 1000 ms. Squares could not touch or overlap each other. Corresponding to the current trial's number of items, randomly selected squares

changed their color to a light grey for 1500 ms sequentially with 1500 ms interstimulus interval. After the last color change, the squares were visible for another 2000 ms before the presentation phase of the Libet clock task started. If the current trial's number of items was zero, a white bar appeared below the rectangular border for 1500 ms to inform the participants of this and after another 1000 ms the Libet clock task started.

After the response phase of the Libet clock task, the array of squares reappeared, and participants had to select the squares that previously changed their color by clicking on them. Participants could freely select and deselect squares and the button to proceed only appeared when the number of selected squares matched the number of items. For comparability with the visual memory task, the sequence in which squares changed color did not matter and participants were informed of this beforehand. There was no time limit for recognition. For trials with zero items, the recognition phase was skipped. Each of the four conditions of the Libet clock task consisted of 6 trials per number of items (0-4), totaling in 120 trials of the spatial memory task. The 30 trials of the memory task per condition of the Libet clock task were presented in a randomized order.

Figure 10

Exemplary sequence of events for the spatial memory task for a trial with three items.



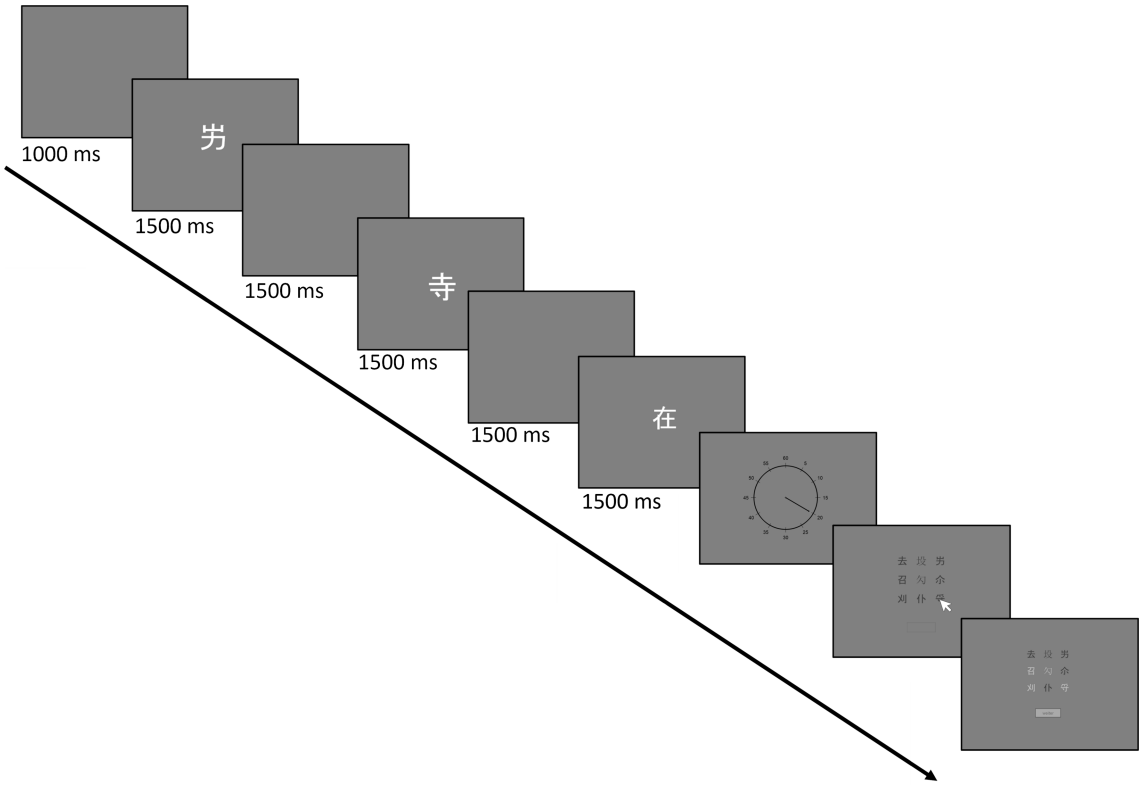
6.1.3.3 Visual Memory Task

The visual memory task was identical to the spatial memory task in timing and sequence of events. Figure 11 shows an exemplary trial. Instead of squares a sequence of zero to four Chinese characters was presented in the center of the screen for 1500 ms each with an interstimulus interval of 1500 ms. Like in the spatial memory task, one trial of the Libet clock task followed, after which participants had to select the presented characters from an array of nine characters including the presented ones and filled up to nine with distractors. Again, there was no time limit for recognizing the characters. In trials with zero items, a white bar appeared for 1500 ms to avoid participants being confused about missing a character. In these trials the recognition phase was skipped.

For each condition, a list of 240 preselected characters was randomized. For each trial, besides the trials with zero items, nine characters were extracted from that list and the targets chosen at random among them. In this way characters did only appear once per condition and in random character combinations across conditions to mitigate familiarity effects.

Figure 11

Exemplary sequence of events for the visual memory task for a trial with three items.



6.1.4 Procedure

Due to time constraints and the length of the experiment, we tested two participants at the same time. They sat on opposite sides of the testing room facing away from each other. Between them was a room divider and they wore noise-cancelling headphones. First, participants read the study information and data protection information before giving digital informed consent. The experimenter informed the participants of the general procedure and the tasks and encouraged them to ask questions. Then they worked through the training trials clearing all ambiguities if any arose. The two instances of the Libet clock task followed and the session ended with the demographic questionnaire. The order of the two instances was counterbalanced across participants. Each instance typically lasted around 45 minutes and the whole session was around two hours. Participants were encouraged to take pauses any time they needed them.

6.1.5 Statistical Analysis

We preregistered analyses for Study 4 (see beginning of the Study 4 section). We calculated Bayesian linear mixed models using the R package brms (Bürkner, 2017) and chose non- or weakly informative priors (Bürkner, 2018; Danböck et al., 2023). For ease of interpretation, we separated the preregistered analyses for Hypothesis 1 and 2 and ran separate models for the Libet clock instances embedded in the spatial and the visual memory task. We also separated the analysis for action and outcome binding, that is the trials in which participants had to attend the key press and the tone respectively. Instead of calculating temporal binding scores as we did in Study 3, we used the judgment errors of the Libet clock task as the outcome and included the predictors of condition (contingent vs. baseline) and number of items (0-4) and their interaction as fixed effects as well as the intercept. The initial maximum model additionally included the random effects for the intercept, condition, number of items and their interaction.

We fitted this model to our data using gaussian, lognormal, skew normal and student's *t*-distributions and identified the most appropriate distribution family with Pareto-*k*s, posterior prediction checks and leave-one-out cross-validation (Gabry et al., 2019; Vehtari et al., 2017). These criteria favored Student's *t*-distribution for all analyses. Following our preregistration, we reduced our model to optimize power (Matuschek et al., 2017). We reduced the random effects by starting with the correlations between random effects followed by the interaction and the main effects. Model reduction was based on leave-one-out cross-validation. We stopped model reduction when the more complex model showed a better fit which was defined as the difference in expected log predictive density being more than two times its *SE*. The random intercept and the fixed effects were not reduced.

Hypothesis 1 predicts an increase in temporal binding with increasing memory load. In terms of judgment errors more items in the memory task should lead to a more positive error in the contingent compared to the baseline condition for action binding and a more negative error in the

contingent compared to the baseline condition for outcome binding. With our coding of the predictor Condition (baseline as the reference condition) this would show in a positive interaction term of condition and number of items for action binding and a negative interaction term for outcome binding. We interpret statistical significance as zero not being included in the 89% CI for two-sided and in the 78% CI for one-sided tests. As Hypothesis 2 postulates the absence of an effect of substantial magnitude, we interpret Bayes factors to interpret evidence for the null hypothesis versus the alternative with Bayes factors between 1/3 and 3 being inconclusive.

Per participant and per condition trials with judgment errors 3.5 *SD* below or above the respective mean were excluded as outlier trials.

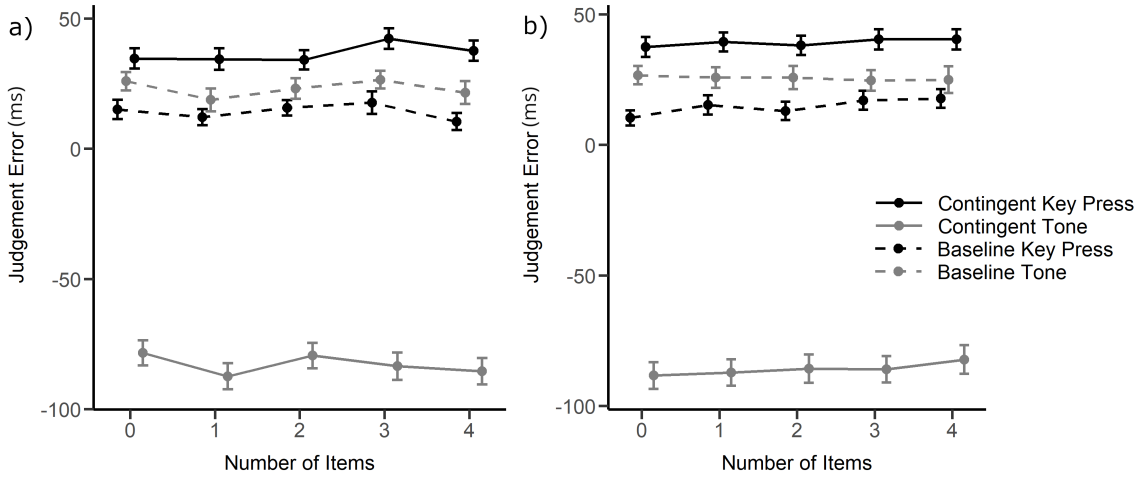
6.2 Results

6.2.1 Memory Task Performance

In the spatial memory task across all four conditions of the Libet clock task, participants correctly recognized 225.0 of 240 targets on average (*SD* = 12.3, range = [174, 239]). In the visual memory task, the average number of correctly recognized targets was 214.6 (*SD* = 23.9, range = [112, 240]). As no participants scored below the chance-level of 80 targets per memory task, no participant was excluded for subpar memory performance. Total number of correctly recognized targets in the two memory tasks correlated at *r* = .59.

Figure 12

Judgment errors in the Libet clock task embedded in the spatial memory task (a) and the visual memory task (b). Temporal binding is visible as the difference between the contingent (solid lines) and baseline conditions (dashed lines) for actions (black) and outcomes (grey). Error bars show ± 1 *SE*.



6.2.2 Action Binding

6.2.2.1 Under Spatial Memory Load

Remaining random effects of the model for judgment errors of the key press during the spatial memory task were the intercept and the factor Condition (see Table 4). The one-sided test for the factor Condition shows that action binding occurs under spatial memory load (B = 25.10, 78%-CI = [18.55, 31.61], $PP_{b>0}$ = 100%, BF_{10} >> 100). On average, the key press is reported as occuring 25 ms later in the contingent compared to the baseline condition (Figure 12a, black lines). Contrary to Hypothesis 1, the interaction of condition and the number of items is not significantly above 0 according to the one-sided test (B = 0.15, 78%-CI = [-0.62, 0.96], $PP_{b>0}$ = 60%, BF_{10} = 1.51). The judgment error does not increase more strongly with increasing spatial memory load in the contingent compared to the baseline condition although the Bayes factor indicates inconclusive evidence. The number of items also does not influence the judgment error across both conditions as shown by the absence of a significant main effect.

Table 4

Bayesian mixed model explaining the judgment error for key presses under spatial memory load with the factor Condition (reference category is the baseline condition) and the continuous predictor Number of Items and their interaction.

			89%-CI			ESS	
Variable	*B*	*SE*	LL	UL	$\widehat{R}$	Bulk	Tail
Fixed Effects							
Intercept	10.62	5.12	2.48	18.80	1.00	3213	5780
Condition	25.10	5.36	16.57	33.60	1.00	5430	9521
n Items	0.17	0.65	-0.86	1.21	1.00	32876	15693
Condition : *n* Items	0.15	0.92	-1.31	1.61	1.00	35357	16254
Random Effects							
SD(Intercept)	46.83	3.64	41.38	52.94	1.00	5095	8250
SD(Condition)	47.06	3.80	41.37	53.42	1.00	9321	14007

Note. *n* Items = number of items, LL = lower limit, UL = upper limit, ESS = effective sample size.

6.2.2.2 Under Visual Memory Load

Remaining random effects of the model for judgment errors of the key press during the visual memory task were the intercept and the factor Condition (see Table 5). Like with spatial memory load, action binding occurs under visual memory load as well (Figure 12b, black lines). The one-sided test of the factor Condition is significant (B = 27.35, 78%-CI = [20.80, 33.93], $PP_{b>0}$ = 100%, BF_{10} >> 100). In accordance with Hypothesis 2, the interaction is non-significant (B = -0.78, 78%-CI = [-1.92, 0.36], $PP_{b>0}$ = 20%, BF_{10} = 0.25) with the Bayes factor indicating substantial evidence for the null hypothesis. Under visual memory load, judgment errors do not increase more strongly with increasing number of items in the contingent compared to the baseline condition. The main effect of number of items is non-significant as well, indicating no influence of visual memory load on judgment errors in both conditions.

Table 5

Bayesian mixed model explaining the judgment error for key presses under visual memory load with the factor Condition (reference category is the baseline condition) and the continuous predictor Number of Items and their interaction.

			89%-CI			ESS	
Variable	*B*	*SE*	LL	UL	$\widehat{R}$	Bulk	Tail
Fixed Effects							
Intercept	9.98	5.15	1.79	18.28	1.00	2403	5246
Condition	27.35	5.33	18.82	35.81	1.00	4348	8538
n Items	0.89	0.65	-0.15	1.93	1.00	33134	15710
Condition : *n* Items	-0.78	0.93	-2.26	0.70	1.00	31293	16738
Random Effects							
SD(Intercept)	48.07	3.68	42.48	54.27	1.00	4648	8014
SD(Condition)	47.20	3.76	41.59	53.53	1.00	7439	11964

Note. *n* Items = number of items, LL = lower limit, UL = upper limit, ESS = effective sample size.

6.2.3 Outcome Binding

6.2.3.1 Under Spatial Memory Load

The random effects for the model for judgment errors for tones under spatial memory load reduced to the intercept and the factor Condition (see Table 6). Outcome binding occurred under spatial memory load with a significant one-sided test of the factor condition (B = -105.98, 78%-CI = [-116.34, -95.46], $PP_{b<0}$ = 100%, BF_{10} >> 100). On average, the tone is reported as occurring 106 ms earlier in the contingent compared to the baseline condition (Figure 12a, grey lines). The interaction of condition and number of items is not significant (B = 0.36, 78%-CI = [-1.02, 1.75], $PP_{b<0}$ = 38%, BF_{10} = 0.61) with the Bayes factor indicating inconclusive evidence. This is contrary to Hypothesis 1. Judgment errors for the tone are not increased more strongly (more negative) with increasing number of items in the contingent compared to the baseline condition. Neither does the number of items influence judgment errors across both conditions indicated by the non-significant main effect.

Table 6

Bayesian mixed model explaining the judgment error for tones under spatial memory load with the factor Condition (reference category is the baseline condition) and the continuous predictor Number of Items and their interaction.

			89%-CI			ESS	
Variable	*B*	*SE*	LL	UL	$\widehat{R}$	Bulk	Tail
Fixed Effects							
Intercept	23.34	5.00	15.45	31.43	1.00	3454	6862
Condition	-105.98	8.57	-119.59	-92.34	1.00	2184	5090
n Items	-1.13	0.79	-2.38	0.14	1.00	30561	15998
Condition : *n* Items	0.36	1.13	-1.46	2.16	1.00	30668	14869
Random Effects							
SD(Intercept)	44.37	3.49	39.18	50.29	1.00	5023	9534
SD(Condition)	79.11	6.03	70.23	89.33	1.00	4156	7404

Note. *n* Items = number of items, LL = lower limit, UL = upper limit, ESS = effective sample size.

6.2.3.2 Under Visual Memory Load

The model for outcome binding embedded in the visual memory task had convergence issues indicated by $\widehat{R} > 1.00$ for some predictors. This problem persisted with 4000 iterations and 1000 warm-up runs. The results in Table 6 are based on 5000 iterations and 1500 warm-up runs which led to satisfactory convergence. The remaining random effects were the intercept and the factor Condition (see Table 7). As with spatial memory load, outcome binding did occur under visual memory load (Figure 12b, grey lines) with a one-sided significant condition effect (B = -115.21, 78%-CI = [-125.73, -104.62], $PP_{b<0}$ = 100%, BF_{10} >> 100). The interaction of condition and number of items was not significant (B = 1.51, 78%-CI = [0.12, 2.89], $PP_{b<0}$ = 9%, BF_{10} = 0.10), confirming Hypothesis 2. In the contingent condition, judgment errors do not become more negative more strongly with increasing number of items compared to the baseline condition. The main effect of number of items is also not significant, showing no influence of memory load on judgment errors across both conditions.

Table 7

Bayesian mixed model explaining the judgment error for tones under visual memory load with the factor Condition (reference category is the baseline condition) and the continuous predictor Number of Items and their interaction.

			89%-CI			ESS	
Variable	*B*	*SE*	LL	UL	$\widehat{R}$	Bulk	Tail
Fixed Effects							
Intercept	25.31	5.85	16.15	34.77	1.00	3492	7455
Condition	-115.21	8.64	-128.85	-101.40	1.00	3525	6605
n Items	-0.05	0.79	-1.32	1.22	1.00	42703	22321
Condition : *n* Items	1.51	1.13	-0.31	3.28	1.00	41031	21654
Random Effects							
SD(Intercept)	53.35	4.21	47.05	60.48	1.00	7179	13315
SD(Condition)	80.14	6.10	70.92	90.46	1.00	5717	11438

Note. *n* Items = number of items, LL = lower limit, UL = upper limit, ESS = effective sample size.

6.2.4 Analysis of Three-Way-Interaction including Memory Type as a Factor

As mentioned in the section on the statistical analysis, the preregistration only included separate two-way-models including Libet clock condition and number of items as predictors. For further insight, we additionally report three-way-models including the memory task (spatial vs. visual) as an additional factor. These results must be seen as exploratory. Going beyond the two-way-models, the three-way-model allows to directly test whether the number of items influences binding differently under spatial and visual memory load. To be more precise, looking at the three-way-interaction we test whether the number of items increases the judgment error more strongly in the contingent condition compared to the baseline condition and whether this differs under the two types of memory load. For action binding, our coding of predictor categories requires a positive regression coefficient of the three-way interaction to indicate a stronger increase of binding with an increasing number of items for the spatial compared to the visual memory task. For outcome binding, this is indicated by a negative regression coefficient.

6.2.4.1 Action Binding

The model for judgment errors of the key press containing memory task as an additional factor retained the random effects of the intercept, condition, number of items, memory task and the interaction of condition and memory task (see Table 8). As with the analyses separated by memory task, the main effect of condition is significant (one-sided, B = 27.17, 78%-CI = [20.76, 33.61], $PP_{b>0}$ = 100%, BF_{10} >> 100). The two-way interaction of condition and number of items is not significant (B = -0.71, 78%-CI = [-1.86, 0.42], $PP_{b>0}$ = 29%, BF_{10} = 0.22). This means that across both memory tasks, action binding does not increase with the number of items. The three-way interaction is also not significant (B = 0.87, 78%-CI = [-0.75, 2.49], $PP_{b>0}$ = 74%, BF_{10} = 2.89). This means that the influence of memory load on action binding does not differ significantly between memory tasks. The Bayes factor favors the alternative hypothesis that there is a three-way interaction but stays just under the preregistered threshold of 3. Descriptively, this result aligns with our hypotheses. However, the Bayes factor indicates inconclusive evidence, and the 78%-CI includes 0.

Table 8

Bayesian mixed model explaining the judgment error for key presses under memory load with the factor Condition (reference is the baseline condition), the continuous predictor Number of Items, the factor Memory Task (reference category is the visual memory task) and their interactions.

			89%-CI			ESS	
Variable	*B*	*SE*	LL	UL	$\widehat{R}$	Bulk	Tail
Fixed Effects							
Intercept	10.27	5.02	2.17	18.10	1.00	1536	2925
Condition	27.17	5.28	18.74	35.45	1.00	2486	4824
n Items	0.82	0.66	-0.24	1.88	1.00	14062	15342
Memory Task	0.31	3.23	-4.85	5.46	1.00	7291	12064
Condition : *n* Items	-0.71	0.93	-2.21	0.76	1.00	13603	14386
Condition : Memory Task	-1.89	4.30	-8.79	4.99	1.00	9121	13080
n Items : Memory Task	-0.65	0.92	-2.13	0.82	1.00	12790	15170
Condition: *n* Items : Memory Task	0.87	1.33	-1.23	2.99	1.00	12247	14640
Random Effects							
SD(Intercept)	47.12	3.57	41.73	53.02	1.00	2889	5950
SD(Condition)	45.53	3.61	40.19	51.64	1.00	5320	9337
SD(*n* Items)	1.16	0.63	0.15	2.16	1.00	3750	5357
SD(Memory Task)	22.18	2.15	18.93	25.76	1.00	7893	12573
SD(Condition : Memory Task)	26.55	2.84	22.28	31.23	1.00	7734	12527

Note. *n* Items = number of items, LL = lower limit, UL = upper limit, ESS = effective sample size.

6.2.4.2 Outcome Binding

The model for outcome binding including memory type as a factor had problems with convergence with $\widehat{R}$ > 1.00 for some predictors. Increasing iterations to 4000 and warm-up runs to 1000 did not alleviate this, so the results in Table 9 were calculated with 5000 iterations and 1500 warm-up runs which yielded satisfactory convergence. Identically to the three-way model for action binding, the three-way model for judgment errors of the tone retained the random effects of the intercept, condition, number of items, memory task and the interaction of condition and memory task (see Table 9). Again, the main effect of condition is significant (one-sided, B = -114.46, 78%-CI = [-125.11, -103.73], $PP_{b<0}$ = 100%, BF_{10} >> 100). The two-way interaction of condition and number of items is not significantly below 0 (one-sided, B = 1.53, 78%-CI = [0.16, 2.9], $PP_{b<0}$ = 9%, BF_{10} = 0.09). According to the 78%-CI, outcome binding decreases

significantly with an increasing number of items across both memory tasks. The three-way interaction is also not significantly below 0 (one-sided, B = -1.06, 78%-CI = [-3.01, 0.86], $PP_{b<0}$ = 75%, BF_{10} = 2.99)

Table 9

Bayesian mixed model explaining the judgment error for tones under memory load with the factor Condition (reference is the baseline condition), the continuous predictor Number of Items, the factor Memory Task (reference category is the visual memory task) and their interactions.

			89%-CI			ESS	
Variable	*B*	*SE*	LL	UL	$\widehat{R}$	Bulk	Tail
Fixed Effects							
Intercept	25.07	5.67	16.16	34.16	1.00	3255	6087
Condition	-114.46	8.75	-128.45	-100.38	1.00	2124	5572
n Items	-0.05	0.80	-1.35	1.23	1.00	20647	21309
Memory Task	-1.30	3.85	-7.46	4.83	1.00	13278	18779
Condition : *n* Items	1.53	1.12	-0.23	3.32	1.00	20314	22055
Condition : Memory Task	8.04	5.65	-1.01	17.06	1.00	11861	17108
n Items : Memory Task	-1.14	1.11	-2.92	0.63	1.00	19635	21896
Condition : *n* Items : Memory Task	-1.06	1.58	-3.58	1.44	1.00	19020	21412
Random Effects							
SD(Intercept)	51.03	3.99	45.01	57.69	1.00	6059	11102
SD(Condition)	79.31	6.08	70.25	89.50	1.00	4689	7774
SD(*n* Items)	1.74	0.80	0.35	2.95	1.00	5488	8331
SD(Memory Task)	26.20	2.64	22.18	30.62	100	12975	17716
SD(Condition : Memory Task)	40.57	3.87	34.72	47.08	1.00	12615	17914

Note. *n* Items = number of items, LL = lower limit, UL = upper limit, ESS = effective sample size.

meaning that the influence of memory load on outcome binding does not differ significantly between memory tasks. As with action binding, the Bayes factor favors the alternative hypothesis that there is a three-way interaction staying barely under the preregistered threshold of 3.

6.3 Preliminary Discussion

The goal of Study 4 was to test the theory that temporal binding in the Libet clock task stems from processes in spatial memory. To test this, we compared binding under increasing levels of spatial memory load with binding under increasing levels of visual memory load. For both action and outcome binding, our preregistered analyses separated by memory type fail to confirm Hypothesis 1. Results suggest that binding does not increase with spatial memory load. However, the Bayes factors

remain inconclusive. The preregistered analyses confirm Hypotheses 2. Bayes factors indicate substantial to strong evidence against the effect of visual memory load on binding for actions and outcomes.

Our additional exploratory analysis added memory type as a factor into the models. Descriptively, the effect of memory load on binding does differ between spatial and visual memory load in the expected direction. Binding is increased more strongly with spatial than visual memory load for both, action and outcome binding. However, for both types of binding the results are non-significant. Although favoring the hypothesis of the presence of this effect, Bayes factors stay just below the threshold of what we preregistered as conclusive evidence.

A limitation of the memory tasks is that they might not bind purely spatial or visual resources respectively. First, we did not use an articulatory suppression task to prevent subvocal verbalization as a memory strategy. This was in part due to testing multiple participants at once. Spatial and visual content might have been remembered partly using phonological resources. Second, we cannot exclude that participants used other strategies to remember the stimuli. For example, they could have imagined shapes between the target squares in the spatial task (e.g., a triangle when the number of targets was three) or they might have relied on local spatial features of the Chinese characters such as the orientation of certain lines that stood out to them. Some participants reported using parts of the characters that resembled familiar objects, such as a house or a Latin letter. These factors might contribute to the overlap in the resources that are bound as indicated by the relatively high correlation between the two tasks.

7 General Discussion

7.1 Integrative Overview

The common aim of the four presented studies was to contribute to the understanding of the current state and potential caveats of research regarding the sense of agency in general and the temporal binding phenomenon in particular. The aim of Study 1 was to test the homogeneity of measured constructs or underlying processes of different temporal binding paradigms. It took a correlational approach and measured temporal binding in two ways, using the Libet clock method and interval estimation in the same group of participants. We adapted the two methods closely from prior research which led the two methods to differ in experimental characteristics like variability of action-outcome delays and the presence of passive action conditions. We elaborated on Study 1 by equalizing these parameters in Study 2. The two studies produce similar results in that temporal binding between the two methods is not correlated positively. This means that binding in the Libet clock method and in interval estimation is based on different intrapersonally stable factors, which suggests different underlying mechanisms.

Study 3 and Study 4 investigated the research question of whether spatial working memory processes are the mechanism behind temporal binding. Study 3 used a correlational design. We tested participants with the Libet clock to measure temporal binding. A 2-back task measuring visual working memory and one measuring spatial working memory were used to differentially measure working memory ability. To control for the confound between the temporal and spatial dimension in the Libet clock task, we conceived a second binding task using directionally randomly appearing dots instead of the continuously moving clock hand. The studies aim was to test whether participants with higher spatial working memory ability showed lower temporal binding scores while visual working memory was assumed to have no such relationship. With the newly conceived binding task we also tested whether we could produce spatial binding between two action- and outcome-associated dots when they are not positioned relative to each other in a temporo-spatially predictable manner. If we

observed such binding, we expected it to be correlated with binding measures in the Libet clock method. While we observed reliable outcome binding in the spatially random binding task, it was uncorrelated to binding in the Libet clock method suggesting different underlying mechanisms. In terms of correlation between temporal binding measures and working memory ability, participants with higher spatial working memory ability as measured by the 2-back task showed lower outcome binding in the Libet clock method. Other binding measures were uncorrelated with working memory ability. Results regarding visual working memory remained inconclusive which prevents us from ruling out a more general, e.g., attentional mechanism influencing outcome binding in the Libet clock method and working memory performance alike.

Study 4 investigated the same theory of spatial working memory processes causing the temporal binding phenomenon using an experimental approach. Participants did the Libet clock binding task in the retention interval of a spatial and a visual working memory task separately. Spatial working memory was taxed by asking participants to remember sequences of squares being highlighted like in the Corsi block tapping test. Visual working memory was loaded with sequences of Chinese characters. Memory load was systematically varied as an experimental factor by varying the number of items. We tested the assumption that binding spatial working memory resources increases binding while visual working memory load does not increase binding. The results did not confirm our theory. Binding did not change with memory load for spatial and visual working memory content. While descriptively pointing in that direction, exploratory analyses could also not conclusively find a stronger influence of spatial working memory load compared to visual working memory load. They also must be interpreted taking their explorative nature into account. This suggests that the theory of binding in the Libet clock method emerging from spatial working memory processes to be inaccurate.

Taken together, the four studies highlight multiple aspects of this field of research in which the underlying mechanisms are not understood sufficiently and where interpretation of results must be done more differentiated manner and with caution. These aspects encompass the independence

of action and outcome binding in the Libet clock on the one hand and of binding phenomena in different binding paradigms on the other hand (first and second study). They also suggest that binding in the Libet clock is not rooted in spatial working memory processes (third and fourth study) although there are clues from exploratory analyses that still hint at the possibility of a role of spatial working memory. Due to the weakness of evidence and the exploratory nature of these results further research is required before drawing any conclusions. In general, these results emphasize the need for further research investigating the underlying mechanisms of binding phenomena.

A consistent pattern across experiments was that action and outcome binding in the Libet clock did not correlate while both show satisfactory reliability. This confirms findings in previous research (Tonn et al., 2021) and emphasize the importance of differentially interpreting results regarding the two kinds of binding and of striving to understand the different underlying perceptual and cognitive processes before making substantial claims about results on a purely phenomenological level.

7.2 Theoretical implications

The presented studies offer multiple theoretical implications. The primary results of the first and second study reveal that theoretical accounts of temporal binding effects must take response modality into consideration. Binding effects were found to differ consistently across participants with some showing reliably higher and others reliably lower binding. This applied for both paradigms, the Libet clock method and interval estimation. Despite this, binding under both methods remained independent from each other. On a theoretical level this implies that the underlying mechanisms seem to be tied closely to the perceptual modality in which participants solve the task of perceiving the action-outcome delay. The Libet clock acts as a temporal-visual anchor while interval estimation offers no such assistance requiring participants to perceive the temporal delay using other temporal cues. Due to the lack of external cues in the experimental setting these cues can be assumed to be internal (Allman et al., 2014). As our results speak against a common underlying mechanism, the binding

effects might occur in earlier perceptual stages of the processing of time instead of more abstract higher levels of processing. Another explanation could be that, at least for the Libet clock method, the binding effect is not temporal at all but rather visuo-spatial in that only the perceived clock hand positions are affected. Our results suggest otherwise at least for working memory processes. A more general role of attentional processes is still open for discussion (Schwarz & Weller, 2022). For interval estimation it seems plausible that the effect occurs in the temporal dimension. However, as we used interval estimation with participants responding by entering the perceived action-outcome delay in milliseconds, the effect might also stem from the transfer between perception and symbolic representation.

As mentioned before, a consistent result across our studies is that action and outcome binding in the Libet clock method appear independent. On the theoretical level this carries an important implication. The theoretical account explaining binding in the context of cue integration assumes that events with a higher, more reliable perceptual precision are weighted more strongly in the establishment of a temporal structure of an event chain (Lush et al., 2019; Wolpe et al., 2013; Yamamoto, 2020). According to this account, events with a higher sensory precision relative to the sensory precision of other events influence perception of temporal relations more strongly. This would mean that in the paradigms used in the presented studies the action and the outcome compete for perceptual weight. A higher relative sensory precision of one would lead to stronger binding of the other event, and vice versa. This would result in a negative correlation of action and outcome binding which is not supported by our results matching results in previous research (Wolpe et al., 2013).

The first two studies presented here show the need to differentiate theoretical accounts of binding effects for the various paradigms and response methods. They suggest a distinctive role of the perceptual modality in which timing cues are available. In the third and fourth study we investigated the new idea that the binding effect in the Libet clock method is not temporal in nature at all. Instead, we proposed a mechanism that is situated in the processes of spatial working memory. As far as the

interpretability of our results allows, such a mechanism is not at play in the constitution of the binding effect. This eliminates one alternative, non-temporal explanation of the binding effect and strengthens the idea of an actual temporal nature of said effect. As mentioned above, other non-temporal, e.g., attentional accounts could still be valid explanations.

7.3 Practical implications

The studies presented aim to investigate basic processes underlying the so-called temporal binding phenomenon and provide information about common methods that are used to study this phenomenon. Being basic research, it allows little conclusion for practical implications. On the contrary, our results show that the different methods to measure the binding phenomenon seem to measure different processes. This emphasizes how lacking our understanding of these processes is and that more research is necessary before these methods can be used, e.g., for diagnostic and other applied purposes.

Special care must be taken investigating conditions like psychopathology regarding altered temporal binding (e.g., Haggard et al., 2003) if those conditions also show alterations in possibly binding-related processes. For example, patients with schizophrenia have been shown to exhibit reduced temporal sensitivity in a passive interval discrimination task compared to healthy controls (Lee et al., 2009). Patients performed even worse the lower their short-term memory performance was. Such alterations in time perception could form a basis for altered temporal binding effects in patients. The fact that they can occur in purely passive tasks however shows the risk of confounding variables when drawing conclusions from temporal binding effects to psychopathology including an altered sense of agency. As long as the processes behind temporal binding are not well understood, its paradigms should not be used, e.g., as diagnostic instruments. As evidenced by our research, this gap in understanding must be filled separately for different temporal binding paradigms.

A similar case can be made for temporal binding research and possible applications which can have legal ramifications. Like decisions on psychopathology, judgements about a person's sense of

agency in the context of criminal offenses can have dire consequences. The notion of an objective measure of whether a person has a diminished sense of agency in a certain situation, e.g., under the influence of substances like alcohol (De Pirro et al., 2020) might be tempting. In our studies, temporal binding, especially measured with the Libet clock method, even shows reliability scores commonly viewed as satisfactory for single case decision making. However, when not knowing how exactly the temporal binding effects come to be, there can be no certainty under which circumstances drawn conclusions hold up and when they might be compromised by factors not yet known.

7.4 Limitations of this research

The most prominent limitation of all presented studies is their samples of healthy university sports students. Besides having reached a high educational level and being homogenous in this regard, all participants can be assumed to be physically active to some extent. A link between a higher sense of agency and movement experience has been suggested (W. Christensen et al., 2015) but this might be limited to already learned action-outcome contexts (Tanaka & Imamizu, 2025). While it was not measured, educational level is correlated to socioeconomic status (Liu et al., 2022) making it probable that our samples do not represent the general public in this regard either but have higher than average socioeconomic status. Another implication of these sample characteristics arises from the circumstance that people with a history of migration have less educational opportunities (Winkler, 2025) making it probable that such groups are systemically underrepresented in the presented research. Participants reporting neurological or psychiatric disorders were excluded from analyses. While this eliminates such disorders as confounding variables, this reduces generalizability to the general population and patient populations in particular.

The largest technical measurement error can be expected to be with the Libet clock task. Precision with stimulus presentation and input modality is limited by the refresh rate of the monitor which was 60 Hz corresponding to a frame rate of 16.7 ms. The clock hand rotates at a period of 2560 ms or 154 frames at this frame rate. This means that the full circle is divided into 154 clock hand

positions which still gives a fluid impression of a rotating motion, but the input precision is limited to steps of approximately 2.3 degrees.

Because the experimental setup (e.g., input method, manner of presentation) of the presented studies is similar to previous research (e.g., Haggard et al., 2002) comparability is high. However, this approach mirrors most voluntary everyday action only in an abstract way. Especially the arbitrary action-outcome pairing of a button press with a tone differs from many real-life action-outcome pairings like tactile object manipulation. Arguably the most important context for the processing of actions and their consequences is social interactions. The commonly used methods which are used in this research as well are far removed from such contexts and no direct conclusions should be drawn from the presented results in this regard.

While the relationship of the interval estimation and Libet clock methods have been investigated thoroughly in the presented studies, the role of spatial working memory in the temporal binding phenomenon could be studied from a few additional points of view. In Study 3 we employed a correlational approach comparing working memory ability with temporal binding scores. We still recommend at least two conceptual reference points using multiple working memory modalities such as spatial and visual. However, working memory ability encompasses multiple dimensions. We used a 2-back task as it mirrors the stimulus density of the Libet clock method most closely. This task puts demands on updating processes as stimuli must be discarded as soon as they become irrelevant. At the same time, maintenance processes are challenged to a lesser degree compared to the usual working memory capacity (Conway et al., 2005). Other working memory tasks could be employed to cover more of the conceptual variety of working memory ability.

A similar point can be made regarding the conceptual distinctiveness of the working memory tasks we used. In both studies regarding this subject, Study 3 and Study 4, we used spatial working memory tasks to measure working memory ability or load working memory capacity to bind specific resources. Both tasks, the two-back task with dots appearing in a grid and the Corsi block-tapping test, require

spatial processing. Yet they still offer the possibility to recall the positions in reference to each other by mentally forming shapes with vertices at the target locations. Participants' use of strategy is difficult to control but explicit instructions could contribute to conceptually cleaner measures of spatial working memory functioning.

7.5 Suggestions for future research

Our research focused on spatial working memory as an alternative explanation for binding effects in the Libet clock task. The results suggest this is not an accurate account for the binding effects commonly observed. Future research could follow a more general attentional approach as employed by Schwarz and Weller (2022). They added an additional baseline condition to the Libet clock paradigm in which the clock hand briefly changed color, and participants had to remember the color of this change. This was supposed to bind attentional resources to a clock hand position that was not highlighted by an action or an outcome. They found binding for button presses and tones towards these positions highlighted by color change comparable to the usual contingent conditions. These findings could be expanded upon by placing additional attentional demands on the Libet clock paradigm via a general load on attentional resources or using specific attentional demands, e.g., by asking participants to remember clock hand positions in a dual-task design. Another direction that could produce valuable insights focusses on the original explanation of genuine temporal processes (Haggard et al., 2002). The visuo-spatial anchor that is provided by the Libet clock could be compared to similar supporting stimuli either in the visual domain (Cavazzana et al., 2014) or in other perceptual modalities such as tactile or auditory stimulation. A higher-level temporal binding mechanism should lead to correlated binding phenomena across modalities while lower-level perceptual processes should lead to uncorrelated binding measures within participants.

One strength of the presented studies is the consistent use of appropriate methods to calculate reliability scores. This allows for a more detailed interpretation of results most apparent in correlational designs and most useful in case of correlational null results. These could occur in the case

of a true absence of a meaningful correlation or when reliability of the two measures obscures a true correlation. Knowing about the reliability of measures in the present sample and under the specific circumstances of data collection allows for more valid conclusions. It is not enough to rely on reliability estimates from previous research, even when experimental procedures are employed unaltered. We derived reliability scores using different methods depending on the outcome measure (Parsons et al., 2019). The simplest case occurs when trials are combined interchangeably to calculate participant scores making internal consistency a suitable reliability indicator. While dominated by the coefficient alpha, better alternatives for this case are readily available (Bentler, 2009; McNeish, 2018). When the measure of interest is a more complex index, the reliability calculation must account for and emulate the calculation from different kinds of trials or even certain pairings of trials. Permutation-based split-half approaches are a useful tool in these cases. We support the view of Parsons et al. (2019) and recommend the calculation of reliability scores for all cognitive tasks in every sample and the consideration of these scores in interpreting results. Specifically, we recommend an approach for calculating reliability scores that closely mirrors the calculation of the indices used. For calculating reliability scores for temporal binding measures, we recommend the procedures described in the presented studies.

The presented studies also start to fill a research gap in that they suggest binding phenomena in Libet clock and interval estimation to be independent. While these two are the most commonly used paradigms, other paradigms have been developed such as stimulus anticipation (Buehner & Humphreys, 2009) or the method of constant stimuli (Nolden et al., 2012). Future research could investigate the uniformity or dissociation of binding effects between a wider array of paradigms. This could allow for a more structured pooling of related effects. An obvious comparison would be interval estimation and interval reproduction. This could narrow the conceptual location of the binding effect to the perceptual or later stages of processing for these methods. Interval reproduction, e.g., pressing a button for the same duration as a perceived delay between two events or pressing it twice with the same delay that separated the two events keeps the response in the same modality as the stimulus to

be judged. However, some translation of the perceived time interval from the perception of external to the reproduction based on internal temporal cues might still occur. Interval estimation adds an additional transformation of the perceived interval into a symbolic representation like a number of milliseconds which involves additional higher-level processes. A direct comparison of these two methods might help to quantify the additionally introduced variance of such an informational translation into a symbolic representation. One could assume that the calculation of the binding score as the difference between an active and a passive action condition like we employed it in Study 1 and Study 2 would cancel out any effect of a translation into the symbolic domain. On the other hand, there could still be differences in this process depending on the absence or availability of internal action cues.

Another implication for future research is the viability of the dual-task approach as employed in Study 4 for investigating mechanisms behind and factors influencing temporal binding measures in a systematically controlled manner. The main strength is that a temporal binding paradigm can be done twice under closely controllable yet different circumstances while keeping the binding paradigm the same. Loading and taxing specific cognitive resources like different memory modalities offers a precise environment to investigate the involvement of those resources in the binding task. While the dual-task approach can be a valuable tool, it presents the challenge of choosing an appropriate task difficulty. The memory tasks used in Study 4 involve a sequence of visual or spatial stimuli that must be recognized from an array of stimuli after the Libet clock task in any order. More taxing working memory tasks could require participants to also remember the correct sequence or even reproduce the sequence backwards. A simpler parameter to increase working memory demands would be to increase the number of items that must be remembered.

In our studies, we either used constant action-outcome delays or presented different delays in a random order. We also used the same outcome, a sine wave tone, in all trials. While this action-outcome sequence is new to participants in early trials, it becomes familiar and predictable over the

course of the experiment. An interesting approach could encompass a learning phase in which participants get accustomed to a specific action-outcome sequence in terms of timing and outcome characteristics and then introduce a change to this sequence. Investigating whether temporal binding changes gradually with growing familiarity and resets under new conditions might give insights into the underlying processes. Such hypothetical adaptation processes could be compared between different temporal binding paradigms. A neighboring train of thought leads to the question of how temporal binding reacts to trial-by-trial feedback about the accurateness of the estimated action-outcome delay. Informing participants about their over- or underestimation of the temporal interval or, in case of the Libet clock method, of the clock hand position and testing their ability to adapt might be another way of comparing methods to measure temporal binding.

7.6 Final conclusion

Taken together the four studies presented here have important implications for the research of temporal binding. They show the importance to carefully differentiate findings and expand knowledge separately for different perceptual modalities and methods. They suggest a heterogeneity of temporal binding effects dependent on the event of interest and the way in which participants respond. Lastly, they prove the feasibility of dual-task approaches in temporal binding research and provide a template for a hierarchical approach which focusses the research effort from correlational to experimental designs.

8 Declarations

8.1 Ethical Standards

All three studies were performed in line with the principles of the Declaration of Helsinki. Participants were informed about the goal of the study and the privacy policy concerning the data. All participants had to give their written informed consent before participating in this study

8.2 Informed Consent to Participate and to Publish

In all studies participants gave their written informed consent for participating and for the publication of anonymized results before participating.

8.3 Open Research Practices

Raw data, analysis scripts and experimental paradigm files of the presented studies have been made openly available.

Study 1 was not preregistered. Code for the experimental paradigms, data and analysis scripts can be found here: https://osf.io/gwjqf/

Study 2 was preregistered (https://osf.io/jt9us). Code for the experimental paradigms, data and analysis scripts are available here: https://osf.io/gwjqf/

Study 3 was preregistered (https://osf.io/6prgd). Code for the experimental paradigms, data and analysis scripts is openly accessible: doi.org/10.17605/OSF.IO/2NM34

Study 4 was preregistered (https://osf.io/vxagf). Code for the experimental paradigms, data and analysis scripts can be retrieved here: doi.org/10.17605/OSF.IO/5FK7Z

8.4 Acknowledgements

I thank the entire research team of the institute of sport science of the university of Regensburg. For all studies they provided helpful discussions. Special thanks go to Dr. Philipp Hofmann, Dr. Leonardo Jost, Dr. Franziska Schroter and Miro Ebert for advice on statistics and experimental design. I also thank Dr. Franziska Schroter for proof-reading. Finally, I thank my supervisor Dr. Petra Jansen for her guidance during this research process. In the midst of creative and progressively more intricate and labyrinthine ideas she has been a voice of practicality and realistic scope of implementation.

8.5 Conflict of Interests

There were no conflicts of interest in all the studies presented.

8.6 Funding

There was no funding in all the studies presented.

9 References

Aarts, H., & van den Bos, K. (2011). On the Foundations of Beliefs in Free Will: Intentional Binding and Unconscious Priming in Self-Agency. *Psychological Science*, *22*(4), 532–537. https://doi.org/10.1177/0956797611399294

Allman, M. J., Teki, S., Griffiths, T. D., & Meck, W. H. (2014). Properties of the Internal Clock: First- and Second-Order Principles of Subjective Time. *Annual Review of Psychology*, *65*(1), 743–771. https://doi.org/10.1146/annurev-psych-010213-115117

Antusch, S., Custers, R., Marien, H., & Aarts, H. (2021). Studying the sense of agency in the absence of motor movement: An investigation into temporal binding of tactile sensations and auditory effects. *Experimental Brain Research*, *239*(6), 1795–1806. https://doi.org/10.1007/s00221-021-06087-8

Aytemur, A., & Levita, L. (2021). A reduction in the implicit sense of agency during adolescence compared to childhood and adulthood. *Consciousness and Cognition*, *87*, 103060. https://doi.org/10.1016/j.concog.2020.103060

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423. https://doi.org/10.1016/S1364-6613(00)01538-2

Baguley, T., & Kaye, D. (2010). Review of Understanding psychology as a science: An introduction to scientific and statistical inference. *British Journal of Mathematical and Statistical Psychology*, *63*(3), 695–698. https://doi.org/10.1348/000711009X481027

Balslev, D., Cole, J., & Miall, R. C. (2007). Proprioception Contributes to the Sense of Agency during Visual Observation of Hand Movements: Evidence from Temporal Judgments of Action. *Journal of Cognitive Neuroscience*, *19*(9), 1535–1541. https://doi.org/10.1162/jocn.2007.19.9.1535

Barlas, Z., Hockley, W. E., & Obhi, S. S. (2017). The effects of freedom of choice in action selection on perceived mental effort and the sense of agency. *Acta Psychologica*, *180*, 122–129. https://doi.org/10.1016/j.actpsy.2017.09.004

Barlas, Z., Hockley, W. E., & Obhi, S. S. (2018). Effects of free choice and outcome valence on the sense of agency: Evidence from measures of intentional binding and feelings of control. *Experimental Brain Research*, *236*(1), 129–139. https://doi.org/10.1007/s00221-017-5112-3

Barlas, Z., & Kopp, S. (2018). Action Choice and Outcome Congruency Independently Affect Intentional Binding and Feeling of Control Judgments. *Frontiers in Human Neuroscience*, *12*, 137. https://doi.org/10.3389/fnhum.2018.00137

Barlas, Z., & Obhi, S. S. (2013). Freedom, choice, and the sense of agency. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00514

Beck, B., Di Costa, S., & Haggard, P. (2017). Having control over the external world increases the implicit sense of agency. *Cognition*, *162*, 54–60. https://doi.org/10.1016/j.cognition.2017.02.002

Bentler, P. M. (2009). Alpha, Dimension-Free, and Model-Based Internal Consistency Reliability. *Psychometrika*, *74*(1), 137–143. https://doi.org/10.1007/s11336-008-9100-1

Blakemore, S.-J., Frith, C. D., & Wolpert, D. M. (1999). Spatio-Temporal Prediction Modulates the Perception of Self-Produced Stimuli. *Journal of Cognitive Neuroscience*, *11*(5), 551–559. https://doi.org/10.1162/089892999563607

Blakemore, S.-J., Wolpert, D. M., & Frith, C. D. (2002). Abnormalities in the awareness of action. *Trends in Cognitive Sciences*, *6*(6), 237–242. https://doi.org/10.1016/S1364-6613(02)01907-1

Blakey, E., Tecwyn, E. C., McCormack, T., Lagnado, D. A., Hoerl, C., Lorimer, S., & Buehner, M. J. (2019). When causality shapes the experience of time: Evidence for temporal binding in young children. *Developmental Science*, *22*(3). https://doi.org/10.1111/desc.12769

Borhani, K., Beck, B., & Haggard, P. (2017). Choosing, Doing, and Controlling: Implicit Sense of Agency Over Somatosensory Events. *Psychological Science*, *28*(7), 882–893. https://doi.org/10.1177/0956797617697693

Braun, N., Thorne, J. D., Hildebrandt, H., & Debener, S. (2014). Interplay of Agency and Ownership: The Intentional Binding and Rubber Hand Illusion Paradigm Combined. *PLoS ONE*, *9*(11), e111967. https://doi.org/10.1371/journal.pone.0111967

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

Buehner, M. J., & Humphreys, G. R. (2009). Causal Binding of Actions to Their Effects. *Psychological Science*, *20*(10), 1221–1228. https://doi.org/10.1111/j.1467-9280.2009.02435.x

Bürkner, P.-C. (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, *80*(1). https://doi.org/10.18637/jss.v080.i01

Bürkner, P.-C. (2018). Advanced Bayesian Multilevel Modeling with the R Package brms. *The R Journal*, *10*(1), 395. https://doi.org/10.32614/RJ-2018-017

Caspar, E. A., Cleeremans, A., & Haggard, P. (2015). The relationship between human agency and embodiment. *Consciousness and Cognition*, *33*, 226–236. https://doi.org/10.1016/j.concog.2015.01.007

Cavazzana, A., Begliomini, C., & Bisiacchi, P. S. (2014). Intentional binding effect in children: Insights from a new paradigm. *Frontiers in Human Neuroscience*, *8*, 651. https://doi.org/10.3389/fnhum.2014.00651

Cavazzana, A., Begliomini, C., & Bisiacchi, P. S. (2017). Intentional binding as a marker of agency across the lifespan. *Consciousness and Cognition*, *52*, 104–114. https://doi.org/10.1016/j.concog.2017.04.016

Cavazzana, A., Penolazzi, B., Begliomini, C., & Bisiacchi, P. S. (2015). Neural underpinnings of the 'agent brain': New evidence from transcranial direct current stimulation. *European Journal of Neuroscience*, *42*(3), 1889–1894. https://doi.org/10.1111/ejn.12937

Chiarella, S. G., Makwana, M., Simione, L., Hartkamp, M., Calabrese, L., Raffone, A., & Srinivasan, N. (2020). Mindfulness Meditation Weakens Attachment to Self: Evidence from a Self vs Other Binding Task. *Mindfulness*, *11*(10), 2411–2422. https://doi.org/10.1007/s12671-020-01457-9

Christensen, J. F., Di Costa, S., Beck, B., & Haggard, P. (2019). I just lost it! Fear and anger reduce the sense of agency: A study using intentional binding. *Experimental Brain Research*, *237*(5), 1205–1212. https://doi.org/10.1007/s00221-018-5461-6

Christensen, J. F., Yoshie, M., Di Costa, S., & Haggard, P. (2016). Emotional valence, sense of agency and responsibility: A study using intentional binding. *Consciousness and Cognition*, *43*, 1–10. https://doi.org/10.1016/j.concog.2016.02.016

Christensen, W., Bicknell, K., McIlwain, D., & Sutton, J. (2015). The sense of agency and its role in strategic control for expert mountain bikers. *Psychology of Consciousness: Theory, Research, and Practice*, *2*(3), 340–353. https://doi.org/10.1037/cns0000066

Colle, L., Hilviu, D., Rossi, R., Garbarini, F., & Fossataro, C. (2020). Self-Harming and Sense of Agency in Patients With Borderline Personality Disorder. *Frontiers in Psychiatry*, *11*, 449. https://doi.org/10.3389/fpsyt.2020.00449

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*(5), 769–786. https://doi.org/10.3758/BF03196772

Cornelio Martinez, P. I., De Pirro, S., Vi, C. T., & Subramanian, S. (2017). Agency in Mid-air Interfaces. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 2426–2439. https://doi.org/10.1145/3025453.3025457

Corsi, P. M. (1972). Human memory and the medial temporal region of the brain. *Dissertation Abstracts International*, *34*(819B).

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281–302. https://doi.org/10.1037/h0040957

Cubillas, C. P., Landáburu, Í., & Matute, H. (2020). Methodological Factors Involved in the Study of Temporal Binding Using the Open Source Software Labclock Web. *Frontiers in Psychology*, *11*, 1040. https://doi.org/10.3389/fpsyg.2020.01040

Danböck, S. K., Franke, L. K., Miedl, S. F., Liedlgruber, M., Bürkner, P.-C., & Wilhelm, F. H. (2023). Experimental induction of peritraumatic dissociation: The role of negative affect and pain and their psychophysiological and neural correlates. *Behaviour Research and Therapy*, *164*, 104289. https://doi.org/10.1016/j.brat.2023.104289

De Pirro, S., Lush, P., Parkinson, J., Duka, T., Critchley, H. D., & Badiani, A. (2020). Effect of alcohol on the sense of agency in healthy humans. *Addiction Biology*, *25*(4). https://doi.org/10.1111/adb.12796

Desantis, A., & Haggard, P. (2016). Action-outcome learning and prediction shape the window of simultaneity of audiovisual outcomes. *Cognition*, *153*, 33–42. https://doi.org/10.1016/j.cognition.2016.03.009

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

Desantis, A., Weiss, C., Schütz-Bosbach, S., & Waszak, F. (2012). Believing and Perceiving: Authorship Belief Modulates Sensory Attenuation. *PLoS ONE*, *7*(5), e37959. https://doi.org/10.1371/journal.pone.0037959

D'Esposito, M., & Postle, B. R. (2015). The Cognitive Neuroscience of Working Memory. *Annual Review of Psychology*, *66*(1), 115–142. https://doi.org/10.1146/annurev-psych-010814-015031

Dewey, J. A., & Knoblich, G. (2014). Do Implicit and Explicit Measures of the Sense of Agency Measure the Same Thing? *PLoS ONE*, *9*(10), e110118. https://doi.org/10.1371/journal.pone.0110118

Di Plinio, S., Arnò, S., Perrucci, M. G., & Ebisch, S. J. H. (2019). Environmental control and psychosis-relevant traits modulate the prospective sense of agency in non-clinical individuals. *Consciousness and Cognition*, *73*, 102776. https://doi.org/10.1016/j.concog.2019.102776

Dienes, Z. (2008). *Understanding psychology as a science: An introduction to scientific and statistical inference*. Macmillan International Higher Education.

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00781

Dienes, Z. (2016). How Bayes factors change scientific practice. *Journal of Mathematical Psychology*, *72*, 78–89. https://doi.org/10.1016/j.jmp.2015.10.003

Dienes, Z. (2019). How do I know what my theory predicts? *Advances in Methods and Practices in Psychological Science*, *2*(4), 364–377. https://doi.org/10.1177/2515245919876960

Dienes, Z. (2021). How to use and report Bayesian hypothesis tests. *Psychology of Consciousness: Theory, Research, and Practice*, *8*(1), 9–26. https://doi.org/10.1037/cns0000258

Dienes, Z. (n.d.). *Making the most of your data with Bayes*. http://www.lifesci.sussex.ac.uk/home/Zoltan_Dienes/inference/Bayes.htm

Dubynin, I. A., Yashin, A. S., Velichkovsky, B. M., & Shishkin, S. L. (2021). An experimental paradigm for studying sense of agency in joint human-machine motor actions. *Experimental Brain Research*, *239*(6), 1951–1961. https://doi.org/10.1007/s00221-021-06105-9

Eagleman, D. M. (2008). Human time perception and its illusions. *Current Opinion in Neurobiology*, *18*(2), 131–136. https://doi.org/10.1016/j.conb.2008.06.002

Eagleman, D. M. (2013). Motor-sensory recalibration modulates perceived simultaneity of cross-modal events at different distances. *Frontiers in Psychology*, 12. https://doi.org/10.3389/fpsyg.2013.00046

Ebert, J. P., & Wegner, D. M. (2010). Time warp: Authorship shapes the perceived timing of actions and events. *Consciousness and Cognition*, *19*(1), 481–489. https://doi.org/10.1016/j.concog.2009.10.002

Ecker, U. K. H., Lewandowsky, S., & Oberauer, K. (2014). Removal of information from working memory: A specific updating process. *Journal of Memory and Language*, *74*, 77–90. https://doi.org/10.1016/j.jml.2013.09.003

Endo, S., Fröhner, J., Musić, S., Hirche, S., & Beckerle, P. (2020). Effect of External Force on Agency in Physical Human-Machine Interaction. *Frontiers in Human Neuroscience*, *14*, 114. https://doi.org/10.3389/fnhum.2020.00114

Engbert, K., Wohlschläger, A., Thomas, R., & Haggard, P. (2007). Agency, subjective time, and other minds. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(6), 1261–1268. https://doi.org/10.1037/0096-1523.33.6.1261

Farrer, C., & Frith, C. D. (2002). Experiencing Oneself vs Another Person as Being the Cause of an Action: The Neural Correlates of the Experience of Agency. *NeuroImage*, *15*(3), 596–603. https://doi.org/10.1006/nimg.2001.1009

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Fereday, R., & Buehner, M. J. (2017). Temporal binding and internal clocks: No evidence for general pacemaker slowing. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(5), 971–985. https://doi.org/10.1037/xhp0000370

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*, 104005. https://doi.org/10.1016/j.cognition.2019.06.017

Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., & Gelman, A. (2019). Visualization in Bayesian Workflow. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *182*(2), 389–402. https://doi.org/10.1111/rssa.12378

Gallagher, S. (2000). Philosophical conceptions of the self: Implications for cognitive science. *Trends in Cognitive Sciences*, *4*(1), 14–21. https://doi.org/10.1016/S1364-6613(99)01417-5

Gentsch, A., & Schütz-Bosbach, S. (2011). I Did It: Unconscious Expectation of Sensory Consequences Modulates the Experience of Self-agency and Its Functional Signature. *Journal of Cognitive Neuroscience*, *23*(12), 3817–3828. https://doi.org/10.1162/jocn_a_00012

Graham-Schmidt, K. T., Martin-Iverson, M. T., Holmes, N. P., & Waters, F. A. V. (2016). When one's sense of agency goes wrong: Absent modulation of time perception by voluntary actions and reduction of perceived length of intervals in passivity symptoms in schizophrenia. *Consciousness and Cognition*, *45*, 9–23. https://doi.org/10.1016/j.concog.2016.08.006

Haatveit, B. C., Sundet, K., Hugdahl, K., Ueland, T., Melle, I., & Andreassen, O. A. (2010). The validity of *d* prime as a working memory index: Results from the "Bergen *n* -back task." *Journal of Clinical and Experimental Neuropsychology*, *32*(8), 871–880. https://doi.org/10.1080/13803391003596421

Haering, C., & Kiesel, A. (2012). Mine is Earlier than Yours: Causal Beliefs Influence the Perceived Time of Action Effects. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00393

Haering, C., & Kiesel, A. (2014). Intentional Binding is independent of the validity of the action effect's identity. *Acta Psychologica*, *152*, 109–119. https://doi.org/10.1016/j.actpsy.2014.07.015

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

Haggard, P., Martin, F., Taylor-Clarke, M., Jeannerod, M., & Franck, N. (2003). Awareness of action in schizophrenia. *NeuroReport*, *14*(7), 1081–1085. https://doi.org/10.1097/01.wnr.0000073684.00308.c0

Hartley, A. A., Speer, N. K., Jonides, J., Reuter-Lorenz, P. A., & Smith, E. E. (2001). Is the dissociability of working memory systems for name identity, visual-object identity, and spatial location maintained in old age? *Neuropsychology*, *15*(1), 3–17. https://doi.org/10.1037/0894-4105.15.1.3

Hascalovitz, A. (Chen), & Obhi, S. S. (2015). Personality and intentional binding: An exploratory study using the narcissistic personality inventory. *Frontiers in Human Neuroscience*, *9*. https://doi.org/10.3389/fnhum.2015.00013

Hoerl, C., Lorimer, S., McCormack, T., Lagnado, D. A., Blakey, E., Tecwyn, E. C., & Buehner, M. J. (2020). Temporal Binding, Causation, and Agency: Developing a New Theoretical Framework. *Cognitive Science*, *44*(5). https://doi.org/10.1111/cogs.12843

Hughes, G., Desantis, A., & Waszak, F. (2013). Mechanisms of intentional binding and sensory attenuation: The role of temporal prediction, temporal control, identity prediction, and motor prediction. *Psychological Bulletin*, *139*(1), 133–151. https://doi.org/10.1037/a0028566

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

Humphreys, G. R., & Buehner, M. J. (2010). Temporal binding of action and effect in interval reproduction. *Experimental Brain Research*, *203*(2), 465–470. https://doi.org/10.1007/s00221-010-2199-1

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

Jeffreys, H. (1931). *Scientific inference*. Cambridge Univ. Press.

Jenkinson, P. M., Edelstyn, N. M. J., Preston, C., & Ellis, S. J. (2015). Anarchic hand with abnormal agency following right inferior parietal lobe damage: A case report. *Neurocase*, *21*(4), 471–478. https://doi.org/10.1080/13554794.2014.925936

Kass, R. E., & Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association*, *90*(430), 773–795. https://doi.org/10.1080/01621459.1995.10476572

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

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

Klauer, K. C., & Zhao, Z. (2004). Double Dissociations in Visual and Spatial Short-Term Memory. *Journal of Experimental Psychology: General*, *133*(3), 355–381. https://doi.org/10.1037/0096-3445.133.3.355

Kumar, N., Manjaly, J. A., & Sunny, M. M. (2015). The relationship between action-effect monitoring and attention capture. *Journal of Experimental Psychology: General*, *144*(1), 18–23. https://doi.org/10.1037/xge0000032

Lee, K.-H., Bhaker, R. S., Mysore, A., Parks, R. W., Birkett, P. B. L., & Woodruff, P. W. R. (2009). Time perception and its neuropsychological correlates in patients with schizophrenia and in healthy volunteers. *Psychiatry Research*, *166*(2–3), 174–183. https://doi.org/10.1016/j.psychres.2008.03.004

Libet, B., Gleason, C. A., & Wright, E. W. (1983). Time of Conscious Intention to Act in Relation to Onset of Cerebral Activity (Readiness-Potential). *Brain*, *106*(3), 623–642. https://doi.org/10.1093/brain/106.3.623

Liu, J., Peng, P., Zhao, B., & Luo, L. (2022). Socioeconomic Status and Academic Achievement in Primary and Secondary Education: A Meta-analytic Review. *Educational Psychology Review*, *34*(4), 2867–2896. https://doi.org/10.1007/s10648-022-09689-y

Lush, P., Parkinson, J., & Dienes, Z. (2016). Illusory Temporal Binding in Meditators. *Mindfulness*, *7*(6), 1416–1422. https://doi.org/10.1007/s12671-016-0583-z

Lush, P., Roseboom, W., Cleeremans, A., Scott, R., 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/https://doi.org/10.1037/xhp0000661

Ly, A., Verhagen, J., & Wagenmakers, E.-J. (2016). Harold Jeffreys's default Bayes factor hypothesis tests: Explanation, extension, and application in psychology. *Journal of Mathematical Psychology*, *72*, 19–32. https://doi.org/10.1016/j.jmp.2015.06.004

Lynn, M. T., Muhle-Karbe, P. S., Aarts, H., & Brass, M. (2014). Priming determinist beliefs diminishes implicit (but not explicit) components of self-agency. *Frontiers in Psychology*, *5*, 1483. https://doi.org/10.3389/fpsyg.2014.01483

Martin, J.-R. (2013). Experiences of activity and causality in schizophrenia: When predictive deficits lead to a retrospective over-binding. *Consciousness and Cognition*, *22*(4), 1361–1374. https://doi.org/10.1016/j.concog.2013.09.003

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315. https://doi.org/10.1016/j.jml.2017.01.001

McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods*, *23*(3), 412–433. https://doi.org/10.1037/met0000144

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The Unity and Diversity of Executive Functions and Their Contributions to Complex "Frontal Lobe" Tasks: A Latent Variable Analysis. *Cognitive Psychology*, *41*(1), 49–100. https://doi.org/10.1006/cogp.1999.0734

Möller, T. J., Braun, N., Thöne, A.-K., Herrmann, C. S., & Philipsen, A. (2020). The Senses of Agency and Ownership in Patients With Borderline Personality Disorder. *Frontiers in Psychiatry*, *11*, 474. https://doi.org/10.3389/fpsyt.2020.00474

Moore, J. W. (2016). What Is the Sense of Agency and Why Does it Matter? *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.01272

Moore, J. W., Cambridge, V. C., Morgan, H., Giorlando, F., Adapa, R., & Fletcher, P. C. (2013). Time, action and psychosis: Using subjective time to investigate the effects of ketamine on sense of agency. *Neuropsychologia*, *51*(2), 377–384. https://doi.org/10.1016/j.neuropsychologia.2012.07.005

Moore, J. W., & Fletcher, P. C. (2012). Sense of agency in health and disease: A review of cue integration approaches. *Consciousness and Cognition*, *21*(1), 59–68. https://doi.org/10.1016/j.concog.2011.08.010

Moore, J. W., & Haggard, P. (2008). Awareness of action: Inference and prediction. *Consciousness and Cognition*, *17*(1), 136–144. https://doi.org/10.1016/j.concog.2006.12.004

Moore, J. W., & Obhi, S. S. (2012). Intentional binding and the sense of agency: A review. *Consciousness and Cognition*, *21*(1), 546–561. https://doi.org/10.1016/j.concog.2011.12.002

Moore, J. W., Turner, D. C., Corlett, P. R., Arana, F. S., Morgan, H. L., Absalom, A. R., Adapa, R., de Wit, S., Everitt, J. C., Gardner, J. M., Pigott, J. S., Haggard, P., & Fletcher, P. C. (2011). Ketamine administration in healthy volunteers reproduces aberrant agency experiences associated with schizophrenia. *Cognitive Neuropsychiatry*, *16*(4), 364–381. https://doi.org/10.1080/13546805.2010.546074

Moreton, J., Callan, M. J., & Hughes, G. (2017). How much does emotional valence of action outcomes affect temporal binding? *Consciousness and Cognition*, *49*, 25–34. https://doi.org/10.1016/j.concog.2016.12.008

Nataraj, R., Sanford, S., Shah, A., & Liu, M. (2020). Agency and Performance of Reach-to-Grasp With Modified Control of a Virtual Hand: Implications for Rehabilitation. *Frontiers in Human Neuroscience*, *14*, 126. https://doi.org/10.3389/fnhum.2020.00126

Nolden, S., Haering, C., & Kiesel, A. (2012). Assessing intentional binding with the method of constant stimuli. *Consciousness and Cognition*, *21*(3), 1176–1185. https://doi.org/10.1016/j.concog.2012.05.003

Oberauer, K., & Lin, H.-Y. (2017). An interference model of visual working memory. *Psychological Review*, *124*(1), 21–59. https://doi.org/10.1037/rev0000044

Oren, E., Eitam, B., & Dar, R. (2019). Intentional binding and obsessive-compulsive tendencies: A dissociation between indirect and direct measures of the sense of agency. *Journal of Obsessive-Compulsive and Related Disorders*, *20*, 59–65. https://doi.org/10.1016/j.jocrd.2017.11.002

Parsons, S., Kruijt, A.-W., & Fox, E. (2019). Psychological Science Needs a Standard Practice of Reporting the Reliability of Cognitive-Behavioral Measurements. *Advances in Methods and Practices in Psychological Science*, *2*(4), 378–395. https://doi.org/10.1177/2515245919879695

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

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

Poonian, S. K., & Cunnington, R. (2013). Intentional binding in self-made and observed actions. *Experimental Brain Research*, *229*(3), 419–427. https://doi.org/10.1007/s00221-013-3505-5

Poonian, S. K., McFadyen, J., Ogden, J., & Cunnington, R. (2015). Implicit Agency in Observed Actions: Evidence for N1 Suppression of Tones Caused by Self-made and Observed Actions. *Journal of Cognitive Neuroscience*, *27*(4), 752–764. https://doi.org/10.1162/jocn_a_00745

R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Ren, M., Xu, J., Li, Y., Wang, M., Georgiev, G., Shen, L., Zhao, J., Cao, Z., Zhang, S., Wang, W., Xu, S., Zhou, Z., Chen, S., Chen, X., Shi, X., Tang, X., & Shan, C. (2023). Neural signatures for the n-back task with different loads: An event-related potential study. *Biological Psychology*, *177*, 108485. https://doi.org/10.1016/j.biopsycho.2023.108485

Render, A., & Jansen, P. (2020). Influence of arousal on intentional binding: Impaired action binding, intact outcome binding. *Attention, Perception, & Psychophysics*. https://doi.org/10.3758/s13414-020-02105-z

Ricciardi, L., Haggard, P., de Boer, L., Sorbera, C., Stenner, M.-P., Morgante, F., & Edwards, M. J. (2017). Acting without being in control: Exploring volition in Parkinson's disease with impulsive compulsive behaviours. *Parkinsonism & Related Disorders*, *40*, 51–57. https://doi.org/10.1016/j.parkreldis.2017.04.011

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

Sahaï, A., Pacherie, E., Grynszpan, O., & Berberian, B. (2017). Predictive Mechanisms Are Not Involved the Same Way during Human-Human vs. Human-Machine Interactions: A Review. *Frontiers in Neurorobotics*, *11*, 52. https://doi.org/10.3389/fnbot.2017.00052

Saito, N., Takahata, K., Murai, T., & Takahashi, H. (2015). Discrepancy between explicit judgement of agency and implicit feeling of agency: Implications for sense of agency and its disorders. *Consciousness and Cognition*, *37*, 1–7. https://doi.org/10.1016/j.concog.2015.07.011

Saito, N., Takahata, K., Yamakado, H., Sawamoto, N., Saito, S., Takahashi, R., Murai, T., & Takahashi, H. (2017). Altered awareness of action in Parkinson's disease: Evaluations by explicit and implicit measures. *Scientific Reports*, *7*(1), 8019. https://doi.org/10.1038/s41598-017-08482-0

Sarma, D., & Srinivasan, N. (2021). Intended emotions influence intentional binding with emotional faces: Larger binding for intended negative emotions. *Consciousness and Cognition*, *92*, 103136. https://doi.org/10.1016/j.concog.2021.103136

Schofield, J. S., Shell, C. E., Thumser, Z. C., Beckler, D. T., Nataraj, R., & Marasco, P. D. (2019). Characterization of the Sense of Agency over the Actions of Neural-machine Interface-operated Prostheses. *Journal of Visualized Experiments*, (143), 58702. https://doi.org/10.3791/58702

Schwarz, K. A., & Weller, L. (2022). Distracted to a fault: Attention, actions, and time perception. *Attention, Perception, & Psychophysics*. https://doi.org/10.3758/s13414-022-02632-x

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*, 102807. https://doi.org/10.1016/j.concog.2019.102807

Siebertz, M., & Jansen, P. (2022). Diverging implicit measurement of sense of agency using interval estimation and Libet clock. *Consciousness and Cognition*, *99*, 103287. https://doi.org/10.1016/j.concog.2022.103287

Siebertz, M., & Jansen, P. (2025). Is the temporal binding effect in the Libet clock-task based in spatial working memory? A correlational and a dual-task approach. *Consciousness and Cognition*, *134*, 103909. https://doi.org/10.1016/j.concog.2025.103909

Smith, E. E., Jonides, J., Koeppe, R. A., Awh, E., Schumacher, E. H., & Minoshima, S. (1995). Spatial versus Object Working Memory: PET Investigations. *Journal of Cognitive Neuroscience*, *7*(3), 337–356. https://doi.org/10.1162/jocn.1995.7.3.337

Straube, B., Schülke, R., Drewing, K., Kircher, T., & van Kemenade, B. M. (2017). Hemispheric differences in the processing of visual consequences of active vs. passive movements: A transcranial direct current stimulation study. *Experimental Brain Research*, *235*(10), 3207–3216. https://doi.org/10.1007/s00221-017-5053-x

Suzuki, K., Lush, P., Seth, A. K., & Roseboom, W. (2019). Intentional Binding Without Intentional Action. *Psychological Science*, *30*(6), 842 – 853.

Synofzik, M., Vosgerau, G., & Newen, A. (2008a). Beyond the comparator model: A multifactorial two-step account of agency. *Consciousness and Cognition*, *17*(1), 219–239. https://doi.org/10.1016/j.concog.2007.03.010

Synofzik, M., Vosgerau, G., & Newen, A. (2008b). I move, therefore I am: A new theoretical framework to investigate agency and ownership. *Consciousness and Cognition*, *17*(2), 411–424. https://doi.org/10.1016/j.concog.2008.03.008

Synofzik, M., Vosgerau, G., & Voss, M. (2013). The experience of agency: An interplay between prediction and postdiction. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00127

Takahata, K., Takahashi, H., Maeda, T., Umeda, S., Suhara, T., Mimura, M., & Kato, M. (2012). It's Not My Fault: Postdictive Modulation of Intentional Binding by Monetary Gains and Losses. *PLoS ONE*, *7*(12), e53421. https://doi.org/10.1371/journal.pone.0053421

Tanaka, T., & Imamizu, H. (2025). Sense of agency for a new motor skill emerges via the formation of a structural internal model. *Communications Psychology*, *3*(1), 70. https://doi.org/10.1038/s44271-025-00240-7

Tanaka, T., & Kawabata, H. (2021). Sense of agency is modulated by interactions between action choice, outcome valence, and predictability. *Current Psychology*, *40*(4), 1795–1806. https://doi.org/10.1007/s12144-018-0121-3

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

Tsakiris, M., & Haggard, P. (2003). Awareness of somatic events associated with a voluntary action. *Experimental Brain Research*, *149*(4), 439–446. https://doi.org/10.1007/s00221-003-1386-8

van Kemenade, B. M., Arikan, B. E., Kircher, T., & Straube, B. (2016). Predicting the sensory consequences of one's own action: First evidence for multisensory facilitation. *Attention, Perception, & Psychophysics*, *78*(8), 2515–2526. https://doi.org/10.3758/s13414-016-1189-1

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*(5), 1413–1432. https://doi.org/10.1007/s11222-016-9696-4

Vergauwe, E., Barrouillet, P., & Camos, V. (2009). Visual and spatial working memory are not that dissociated after all: A time-based resource-sharing account. *Journal of Experimental*

Psychology: Learning, Memory, and Cognition, *35*(4), 1012–1028. https://doi.org/10.1037/a0015859

Villa, R., Tidoni, E., Porciello, G., & Aglioti, S. M. (2020). Freedom to act enhances the sense of agency, while movement and goal-related prediction errors reduce it. *Psychological Research*. https://doi.org/10.1007/s00426-020-01319-y

Vinding, M. C., Jensen, M., & Overgaard, M. (2015). The time between intention and action affects the experience of action. *Frontiers in Human Neuroscience*, *9*. https://doi.org/10.3389/fnhum.2015.00366

Vinding, M. C., Pedersen, M. N., & Overgaard, M. (2013). Unravelling intention: Distal intentions increase the subjective sense of agency. *Consciousness and Cognition*, *22*(3), 810–815. https://doi.org/10.1016/j.concog.2013.05.003

Voss, M., Moore, J. W., Hauser, M., Gallinat, J., Heinz, A., & Haggard, P. (2010). Altered awareness of action in schizophrenia: A specific deficit in predicting action consequences. *Brain*, *133*(10), 3104–3112. https://doi.org/10.1093/brain/awq152

Vuorre, M., & Metcalfe, J. (2017). Voluntary action alters the perception of visual illusions. *Attention, Perception, & Psychophysics*, *79*(5), 1495–1505. https://doi.org/10.3758/s13414-017-1321-x

Waszak, F., Cardoso-Leite, P., & Hughes, G. (2012). Action effect anticipation: Neurophysiological basis and functional consequences. *Neuroscience & Biobehavioral Reviews*, *36*(2), 943–959. https://doi.org/10.1016/j.neubiorev.2011.11.004

Wegner, D. M. (2003). The mind's best trick: How we experience conscious will. *Trends in Cognitive Sciences*, *7*(2), 65–69. https://doi.org/10.1016/S1364-6613(03)00002-0

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

Wen, W., Yamashita, A., & Asama, H. (2017). The influence of performance on action-effect integration in sense of agency. *Consciousness and Cognition*, *53*, 89–98. https://doi.org/10.1016/j.concog.2017.06.008

Winkler, O. (2025). Flucht, Bildung und Integration. Zum Stand der Bildungsintegration von geflüchteten Kindern und Jugendlichen in Deutschland. In O. Winkler, *Ungleichheit und Raum* (pp. 315–334). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-49186-4_10

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

Wolpert, D. M., & Kawato, M. (1998). Multiple paired forward and inverse models for motor control. *Neural Networks*, *11*(7–8), 1317–1329. https://doi.org/10.1016/S0893-6080(98)00066-5

Yamamoto, K. (2020). Cue integration as a common mechanism for action and outcome bindings. *Cognition*, *205*, 104423. https://doi.org/10.1016/j.cognition.2020.104423

Yoshie, M., & Haggard, P. (2013). Negative Emotional Outcomes Attenuate Sense of Agency over Voluntary Actions. *Current Biology*, *23*(20), 2028–2032. https://doi.org/10.1016/j.cub.2013.08.034

Yoshie, M., & Haggard, P. (2017). Effects of emotional valence on sense of agency require a predictive model. *Scientific Reports*, *7*(1), 8733. https://doi.org/10.1038/s41598-017-08803-3

Zalla, T., & Sperduti, M. (2015). The sense of agency in autism spectrum disorders: A dissociation between prospective and retrospective mechanisms? *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01278