Binding Processes in Episodic Memory: Measurement, Structure, and Moderators

Marcel R. Schreiner

Inaugural Disseration

Submitted in partial fulfillment of the requirements for the degree of Doctor of Social Sciences in the DFG Research Training Group "Statistical Modeling in Psychology" at the University of Mannheim

1st Supervisor:
Prof. Dr. Thorsten Meiser

2nd Supervisor:
Prof. Dr. Arndt Bröder

Dean of the School of Social Sciences:
Prof. Dr. Michael Diehl

Thesis Evaluators:
Prof. Dr. Edgar Erdfelder
Prof. Dr. Karl Christoph Klauer

Thesis Defense:
January 30, 2023

For my family

Contents

Summary

Episodic memory enables people to remember personally experienced events. While these events consist of different elements, people are able to form coherent memory representations. This requires that an event's constituent elements are bound together in memory. Despite the importance of these binding processes for episodic memory, they are still only poorly understood and our abilities to measure them are limited.

In this thesis, comprising three articles, I provide a new approach for measuring binding effects and use this measure to probe properties of binding processes in episodic memory. In the first article, I introduce the new measurement approach and evaluate its suitability for measuring binding effects in comparison to previous approaches. I show that the approach has good measurement properties and is better suited for measuring binding effects than previous approaches. In the second article, I examine the structure in which event elements are bound together and whether animacy influences binding processes. I show that different binding structures are possible, such as an integrated binding structure, in which event elements are bound into a unitary representation, and a hierarchical binding structure, in which event elements are preferentially bound to particular types of elements. These may lie on a continuum of memory representations with varying degrees of integration. I further show that the presence of an animate element in an event facilitates binding, enabling more coherent memory representations with a higher degree of integration. In addition, awareness regarding commonalities of types of event elements across events may facilitate binding. In the third article, I examine whether agency influences binding processes. I show that the presence of an agentic element in an event may facilitate binding, but evidence was not conclusive and effects may have been concealed due to low memory performance. Agency may thus underlie the previously found facilitating effect of animacy on binding, since animate elements may exert their influence by providing a potential agent in an event.

One aim of my thesis is to provide a new tool for investigating binding processes in episodic memory. An additional aim is to extend our current understanding of binding structures that link together the elements of an event, as well as the factors that moderate binding processes. In doing so, I hope to advance our understanding of binding processes and enable and inform future exploration, as well as theory development and refinement, of this fundamental property underlying episodic memory.

Articles

This cumulative thesis is based on three articles, two of which have been published and one that has been submitted for publication.

Article I

Schreiner, M. R., & Meiser, T. (2022). Measuring binding effects in event-based episodic representations. *Behavior Research Methods*. Advance online publication. https://doi.org/10.3758/s13428-021-01769-1

Article II

Schreiner, M. R., Meiser, T., & Bröder, A. (2022). The binding structure of event elements in episodic memory and the role of animacy. *Quarterly Journal of Experimental Psychology*. Advance online publication. https://doi.org/10.1177/17470218221096148

Article III

Schreiner, M. R., Bröder, A., & Meiser, T. (2022). *Agency effects on the binding of event elements in episodic memory.* Manuscript submitted for publication.

1 Binding Processes in Episodic Memory

In our everyday life we encounter a multitude of events, such as buying a bread at a bakery, meeting a friend in the streets, or having a meeting at work. Episodic memory refers to the capacity of remembering such experiences and has been described as a form of "mental time travel" that allows individuals to re-experience past events (Tulving, 1972, 1983, 1993). These past events are comprised of several different elements. For example, the event of buying bread at a bakery may consist of the vendor selling the bread (a person), the bought bread (an object), and the bakery (a location). In addition, there may be further sensory elements such as the smell of the bread or the noise of an oven. Yet, we do not have isolated representations of these different event elements, but are instead able to remember the entire event in a coherent manner. This necessitates that event elements, despite being represented in different neocortical regions (Alvarez & Squire, 1994; Horner et al., 2015), are bound together in memory to enable the formation of coherent memory representations. The ability to form such bindings develops from early childhood to young adulthood (Ngo et al., 2019; Reese et al., 2011; Schlichting et al., 2017), but decreases in old age (Naveh-Benjamin, 2000; Ngo & Newcombe, 2021; Old & Naveh-Benjamin, 2008). In the present thesis, I investigate a number of fundamental properties of this essential ability underlying episodic memory using statistical modeling.

1.1 Item- vs. Event-Based Representations

Episodic memories may vary in complexity. Some representations may only consist of a single element with specific features, for example an object with a certain color and shape such as the bread one bought at a bakery . Such *item-based representations* are static (see Hunt & Einstein, 1981). More complex episodic memories may incorporate several event elements that can potentially interact (e.g., buying bread at a bakery involves the vendor interacting with the bread). Such *event-based representations* are thus potentially dynamic (see also Rubin & Umanath, 2015). Event-based representations can be considered to be comprised of item-based representations, with storage occurring in a hierarchical manner (see Andermane et al., 2021). Event- and item-based representations may also be distinguished based on the specificity of the stored information, with item-based representations containing more specific information than event-based representations (Hunt & Einstein, 1981). Further, unlike item-based representations, event-based representations incorporate a spatiotemporal context (e.g., Andermane et al., 2021) and allow for the construction of scenes (Robin, 2018; Rubin & Umanath, 2015). This does not necessitate that the specific features of an event's constituent elements, which are stored as item-based representations, are exactly remembered (Rubin & Umanath, 2015). The present thesis focuses on event-based representations.

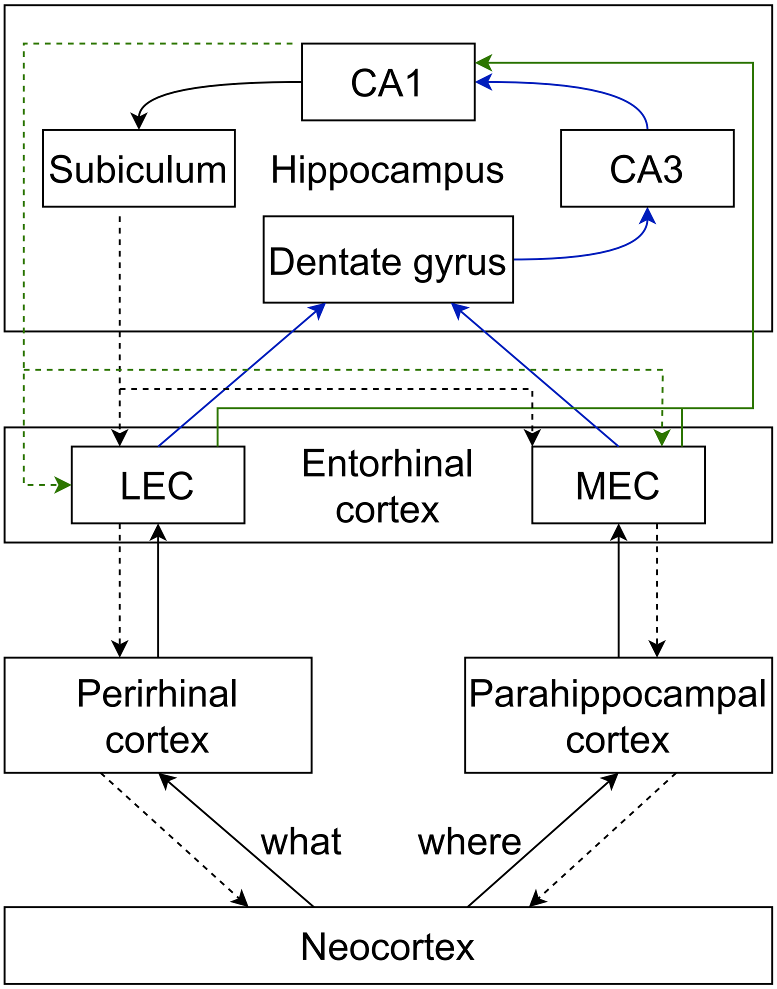
1.2 Neural Foundations of Binding in Episodic Memory

The conceptual distinction between item-based representations and event-based representations, which include spatiotemporal and relational information, is sustained by various neurocognitive mechanisms and brain regions. Generally, the medial temporal lobe, encompassing the perirhinal, entorhinal, and parahippocampal cortex and the hippocampus, plays a crucial role in episodic memory (Eichenbaum et al., 2007; Squire & Zola-Morgan, 1991). Specifically, the formation of event-based representations is primarily supported by the hippocampus, which is commonly seen as the structure responsible for the binding of event elements (Backus et al., 2016; N. J. Cohen & Eichenbaum, 1993; Davachi et al., 2003; Diana et al., 2010; Eichenbaum et al., 2007) and capable of combining different types of information (object, spatial, and temporal information; Sugar & Moser, 2019).

The flow of information between neocortical areas to the hippocampus is depicted in Figure 1. Information from perceptual processing areas in the neocortex is increasingly aggregated via feedforward projections to the hippocampus (Eichenbaum et al., 2007; Rolls, 2016; van Strien et al., 2009). While the processing and binding of temporal information ("when" information) may consist of a widespread brain system involving activity in the hippocampus, lateral entorhinal cortex (LEC), medial enthorhinal cortex (MEC), and additional brain regions (see Eichenbaum, 2017), it has been proposed that there are two pathways for object and spatial information, respectively, which converge in the hippocampus (Eichenbaum et al., 2007; Rolls, 2016; van Strien et al., 2009). A pathway for object information ("what" pathway) involves the perirhinal cortex, which projects to the LEC. A pathway for spatial information ("where" pathway) involves the parahippocampal cortex, which projects to the MEC. Both the LEC and MEC have reciprocal connections with the hippocampus. The hippocampus itself encompasses different subregions — dentate gyrus, cornu ammonis (CA), which can be divided into further subregions (CA1-4), and subiculum (Aggleton & Brown, 1999; Saunders & Rosene, 1988). Within the hippocampus, there are two main pathways, which are strongly involved in binding in episodic memory.

Figure 1

Feedforward projections from the neocortex to the hippocampus via the perirhinal, parahippocampal, and entorhinal cortex (solid lines), and backprojections from hippocampal subfields CA1 and subiculum to the neocortex (dashed lines). Blue arrows make up the trisynapic pathway and green arrows make up the monosynaptic pathway.



Note. what = object information, where = spatial information, LEC = lateral entorhinal cortex, MEC = medial entorhinal cortex, CA = cornu ammonis.

The trisynaptic pathway allows binding within events to form new episodic representations. It involves projections from the entorhinal cortex to the dentate gyrus, which is associated with pattern separation, segregating similar memories by reducing their representational overlap (Bakker et al., 2008; Neunuebel & Knierim, 2014; Rolls, 2016). Dentate gyrus projects to CA3, which is associated with relational binding (i.e., associating individual elements separated by space or time) and may act as an autoassociative network, enabling pattern completion — a whole representation being retrieved by partial activation of the representation through a recall cue (Nakazawa et al., 2002; Neunuebel & Knierim, 2014; Rolls, 2016). CA3 then projects to CA1, which is also associated with relational binding and may be particularly important for binding across temporarily divided encoding episodes (Rolls, 2016; Schlichting et al., 2014). CA1 projects back to the entorhinal cortex and, in addition, to the subiculum, which is the major output structure of the hippocampus (O'Mara et al., 2001; O'Mara, 2005; Rolls, 2016).

The monosyaptic pathway allows learning of regularities and changes across different encoding episodes and events (McClelland et al., 1995; Schapiro et al., 2017; van Strien et al., 2009). It involves direct reciprocal projections between the entorhinal cortex and CA1. Information is retrieved via backprojections from the hippocampus (CA1 and subiculum) to the neocortex (Eichenbaum et al., 2007; Rolls, 2016; van Strien et al., 2009).

In the bakery example, activation regarding the vendor and bread would travel through the "what" pathway and activation regarding the bakery would travel through the "where" pathway. The activation regarding the different types of information would converge in the hippocampus and the different event elements would be bound into a coherent memory representation via the trisynaptic pathway. The dentate gyrus would differentiate event elements in this particular event from other similar events, for example another occasion where one bought a different bread at the same bakery. CA3 and CA1 would then bind the vendor, the bread, and the bakery together, potentially including temporal information such as the time of day one went to the bakery. Then, a coherent memory representation could be retrieved through backprojections from CA1 and via the subiculum to the neocortex.

While investigating the neural mechanisms underlying binding processes in episodic memory is important for a deeper understanding of these processes, binding effects can not only be assessed with neural data, but also with behavioral data. An advantage of assessing binding effects using behavioral instead of neural data is that such studies are much easier to conduct, less resource-intensive, and allow for larger sample sizes. In addition, some aspects of binding may not be properly examinable using neural data, but warrant investigation on a more functional level, which can be achieved by statistical modeling. In the present thesis, I use a theory-driven combination of behavioral experiments and statistical modeling to investigate binding processes in episodic memory.

1.3 Stochastic Dependency of the Retrieval of Event Elements

As a consequence of successful binding, the likelihood of retrieving event elements (e.g., the bread) is increased if other event elements (e.g., the vendor or the bakery) are successfully retrieved, leading to a stochastic dependency of the retrieval of event elements (Arnold et al., 2019; Boywitt & Meiser, 2012a, 2012b; Horner et al., 2015; Horner & Burgess, 2013, 2014; Meiser & Bröder, 2002; Starns & Hicks, 2005, 2008). This stochastic dependency can be viewed as a manifestation of binding processes in episodic memory. Thus, by modeling this dependency, one can draw inferences on binding processes based on behavioral data.

In this thesis, I subsequently introduce a number of existing, contingency-based, approaches for modeling the stochastic dependency of the retrieval of event elements before introducing a newly developed approach (Schreiner & Meiser, 2022; Schreiner, Meiser, & Bröder, 2022) based on item response theory (IRT, Lord, 1980; Lord & Novick, 1968) and evaluating the different approaches regarding their suitability for measuring binding effects in episodic memory. Furthermore, I examine the structure in which different event elements are bound together, and influences of animacy and agency on the binding of event elements in episodic memory. In the first article (Schreiner & Meiser, 2022), I evaluate the different modeling approaches and show that the newly developed IRT-based approach performs best, yielding unbiased estimates, good maintenance of Type I error rates, and high power for detecting binding effects. In the second article (Schreiner, Meiser, & Bröder, 2022), I use this IRT-based approach to investigate the structure in which different event elements are bound together and to examine influences of animacy on binding processes. There, I demonstrate that binding structures may vary, with the possibility of both hierarchical binding structures, in which event elements are preferentially bound to particular types of event elements, and integrated binding structures, in which event elements are bound into a fully integrated representation or engram (cf. Tulving, 1983). I further demonstrate that the presence of an animate element in an event facilitates binding. Finally, in the third article (Schreiner, Bröder, & Meiser, 2022), I examine influences of agency on binding processes, since agency may be a more proximate explanation for the previously found effects of animacy. There, I provide suggestive evidence that the presence of an agentic element in an event facilitates binding.

2 Approaches for Measuring Binding Effects

2.1 Contingency-Based Approaches

One class of approaches for modeling stochastic dependencies of the retrieval of event elements can be applied to data from which one can create meaningful dependency pairs, such as cued recognition (or paired associates matching) and cued recall tasks. Dependency pairs can be, for example, items (i.e., test trials in a memory test) that share a common cue or target element (see Horner & Burgess, 2013). For example, when an event consists of three elements — a person, an object, and a location, as is the case in the bakery example — the cue-target pairs, person–object and person–location, can be considered a dependency pair, since both items share a common cue (the person element). For a given dependency pair *jj'*, one can create a contingency table **X** for each person *i* and event *t*, that shows whether the targets of the dependency pair were retrieved successfully (denoted by 1) or were not retrieved (denoted by 0):

$$\mathbf{X}_{\mathrm{it}}^{\mathrm{jj'}} = \begin{bmatrix} j = 1, j' = 1 & j = 1, j' = 0 \\ j = 0, j' = 1 & j = 0, j' = 0 \end{bmatrix} \qquad (1)$$

By summing these contingency tables over events one obtains a contingency table for each person and dependency pair:

$$\mathbf{X}_{\mathrm{i}}^{\mathrm{jj'}} = \begin{bmatrix} n_{11} & n_{10} \\ n_{01} & n_{00} \end{bmatrix}, \qquad (2)$$

where n_{11} denotes the frequency of the targets of both items *j* and *j'* being successfully retrieved across events, n_{10} denotes the frequency of the target of item *j* being successfully retrieved and that of item *j* not being retrieved, n_{10} denotes the frequency of the target of item *j* not being retrieved and that of item *j'* being successfully retrieved, and n_{00} denotes the frequency of the targets of both items *j* and *j'* not being retrieved. The subsequently described contingency-based approaches for modeling stochastic dependencies of the retrieval of event elements are based on the contingency tables in Equation 2.

2.1.1 Approach by Horner and Burgess

In the approach by Horner and Burgess (2013), two dependency indices are calculated from the contingency tables in Equation 2. First, a data-based dependency index ($D^{\text{data}}_{\text{HB, i}}$), that reflects the average proportion of items in an event whose targets were both successfully retrieved or not retrieved, is calculated by summing the leading diagonal cells of each contingency table per person and dependency pair and dividing this sum by the total number of events[1] T. Then, the index is obtained by averaging across the set of dependency pairs J:

$$D^{\text{data}}_{\text{HB, i}} = \frac{1}{|J|} \sum_{jj\text{'}\in J} \frac{n_{11} + n_{00}}{T} \tag{3}$$

Second, a dependency index from an "independent model" ($D^{\text{ind}}_{\text{HB, i}}$) is calculated by multiplying the probability of successfully retrieving or not retrieving the items' targets as if item responses were independent:

$$D^{\text{ind}}_{\text{HB, i}} = \frac{1}{|J|} \sum_{jj\text{'}\in J} \left(\frac{n_{11} + n_{10}}{T} \frac{n_{11} + n_{01}}{T} + \left(1 - \frac{n_{11} + n_{10}}{T}\right)\left(1 - \frac{n_{11} + n_{01}}{T}\right)\right) \tag{4}$$

The actual dependency measure ($D_{\text{HB, i}}$) is then computed by subtracting the two indices:

$$D_{\text{HB, i}} = D^{\text{data}}_{\text{HB, i}} - D^{\text{ind}}_{\text{HB, i}} \tag{5}$$

This is done to avoid scaling with memory performance, since $D^{\text{data}}_{\text{HB, i}}$ necessarily increases if many or few event elements were successfully retrieved due to strong or poor overall memory performance. The measure can take values between -1 and 1, where 0 indicates independence, positive values indicate dependency (i.e., the likelihood of retrieving an event element is larger if another event element was successfully retrieved), and negative values indicate negative dependency (i.e., the likelihood of retrieving an event element is smaller if another event elements was successfully retrieved).

2.1.2 Yule's Q

Another approach for modeling stochastic dependencies of the retrieval of event elements is Yule's Q (Yule, 1912; cf. Horner & Burgess, 2014; see also Hayman &

---

[1] An event can be broadly described as a set of elements or stimuli that are somehow related (e.g., due to spatial or temporal contiguity).

Tulving, 1989), a standardized odds ratio which is commonly used as a measure of association in memory research (e.g., Kahana, 2002; Kahana et al., 2005). Yule's Q can be calculated from the contingency tables in Equation 2 to receive person- and dependency-pair-specific indices using the following equation:

$$Q_{\mathrm{i}}^{\mathrm{jj'}} = \frac{n_{11}n_{00} - n_{10}n_{01}}{n_{11}n_{00} + n_{10}n_{01}} \tag{6}$$

By averaging across dependency pairs one can then receive person-specific indices:

$$Q_{\mathrm{i}} = \frac{1}{|J|} \sum_{jj' \in J} Q_{\mathrm{i}}^{\mathrm{jj'}} \tag{7}$$

The interpretation of this measure is equivalent to the interpretation of the measure by Horner and Burgess (2013).

2.1.3 Adjusted Yule's Q

A problem of Yule's Q is that zero frequencies in the contingency table it is calculated from (e.g., $n_{10} = 0$) cause it to be bound at its extreme values (-1 or 1) or become undefined. Consider the contingency table in Table 1 from an example by Hintzman (1980) with two outcomes (Y_1 and Y_2).

Table 1

Example contingency table with two outcomes Y_1 and Y_2 from an example by Hintzman (1980).

	$Y_2 = 1$	$Y_2 = 0$
$Y_1 = 1$	20	0
$Y_1 = 0$	60	20

Calculating Yule's Q from Equation 6 results in a value of $Q = 1$ (indicating a perfect positive association between the two outcomes). One can circumvent this problem by adding a constant c to each cell of the contingency tables in Equation 2 (cf. Burton et al., 2019; Horner & Burgess, 2014; see also Snodgrass & Corwin, 1988) to calculate an adjusted Yule's Q (Q_{a}) using Equations 6 and 7. Adding $c = 0.5$ to each cell of the contingency table in the example in Table 1 results in a value of $Q_{\mathrm{a}} = 0.87$.

2.1.4 Simpson's Paradox

A general problem of the contingency-based approaches is that they rely on contingency tables that are aggregated across events. This makes them prone to Simpson's paradox (Hintzman, 1972, 1980; Simpson, 1951) — if 2 × 2 contingency tables are collapsed into a summary one, the relationship of the two outcomes may differ from the one in the original tables. For instance, consider the example by Hintzman (1980) depicted in Table 2. Both of the original tables (A and B) yield values of $Q_\mathrm{a} = 0.87$ (using $c = 0.5$), indicating a positive relationship of the two outcomes. However, there is a third unobserved variable Z that is positively related to the first outcome and negatively related to the second outcome of the two contingency tables (Hintzman, 1980). The first original table corresponds to $Z = 1$ and second original table corresponds to $Z = 2$. Collapsing the two contingency tables (i.e., collapsing over Z) results in the summary table ($A + B$). The summary table yields a value of Q_a = -0.38, indicating a negative relationship of the two outcomes (Y_1 and Y_2). Thus, the relationship of the two outcomes is inverted compared to the relationship in the original tables. Simpson's paradox can occur due to confounding with person differences, item differences, or person-item interactions (Hintzman, 1972, 1980; see also Burton et al., 2017). Since all of the presented contingency-based approaches yield person-specific dependency estimates, confounding with person differences is not an issue. However, the approaches may be subject to confounding with item differences and person-item interactions.

Table 2

Example by Hintzman (1980) for collapsing two contingency tables into a summary one and associated Q_a values.

A			B			$A + B$		
$Z = 1$	$Y_2 = 1$	$Y_2 = 0$	$Z = 2$	$Y_2 = 1$	$Y_2 = 0$		$Y_2 = 1$	$Y_2 = 0$
$Y_1 = 1$	20	0	$Y_1 = 1$	20	60	$Y_1 = 1$	40	60
$Y_1 = 0$	60	20	$Y_1 = 0$	0	20	$Y_1 = 0$	60	40
$Q_\mathrm{a} = 0.87$			$Q_\mathrm{a} = 0.87$			$Q_\mathrm{a} = -0.38$		

Note. Q_a were calculated by adding $c = 0.5$ to each cell of the contingency tables.

2.2 An IRT-Based Approach

Schreiner, M. R., & Meiser, T. (2022). Measuring binding effects in event-based episodic representations. *Behavior Research Methods*. Advance online publication. https://doi.org/10.3758/s13428-021-01769-1

2.2.1 Parametric Variant

In the first article (Schreiner & Meiser, 2022; see also the second article, Schreiner, Meiser, & Bröder, 2022), we introduced a novel approach for measuring binding effects by modeling the stochastic dependency of the retrieval of event elements that is based on item response theory (IRT; Lord, 1980; Lord & Novick, 1968). This approach takes individual item responses rather than aggregated contingency tables as input and is consequently not prone to confounding due to Simpson's paradox. Since IRT takes person and item differences, and person-item interactions into account, confounding with these covariates is avoided. The approach is based on the three-parameter logistic IRT model by Birnbaum (1968), because this model allows one to take guessing into account, which may frequently occur in memory tests (see e.g., Huff et al., 2011). Thus, one can model the probability of person i to give a correct response u to item j given a latent trait θ (i.e., memory performance in the current model application), item difficulties β, item discrimination parameters α, and item guessing parameters γ:

$$P(u_{ij} = 1) = \gamma_j + (1 - \gamma_j)\frac{e^{\alpha_j(\theta_i - \beta_j)}}{1 + e^{\alpha_j(\theta_i - \beta_j)}} \tag{8}$$

For the purpose of measuring binding effects, this model can usually be simplified in practical applications. In experimental investigations of binding processes, events are usually randomly generated for each participant. Thus, one may fix the item discrimination parameters to 1 (cf. the Rasch model; Rasch, 1960). If testing involves several response alternatives one may a priori fix the guessing parameters to a constant g, for example equal to the stochastic guessing probability given a fixed number of response alternatives (e.g., 0.2 for five response alternatives). Such a simplified model is described by the following equation:

$$P(u_{ij} = 1) = g + (1 - g)\frac{e^{\theta_i - \beta_j}}{1 + e^{\theta_i - \beta_j}} \tag{9}$$

The approach utilizes violations of an assumption inherent in many IRT models that follow from successful binding of event elements. The assumption of local independence (LI) implies that item responses are independent after partialing out the latent trait (de Ayala, 2009; Lazarsfeld & Henry, 1968) and that item residual correlations are thus zero. However, given successful binding, there are additional event-specific effects that lead to a violation of the LI assumption. This leads to item residual correlations that systematically deviate from zero, such that elements of a common event are more likely to be retrieved together, or not to be retrieved together, than elements of different events. In the current approach, item residual correlations are estimated using the Q_3 statistic (Yen, 1984). This statistic can be calculated for each item pair *jj'* (i.e., for each pairwise combination of test trials in a memory test) in four steps: (1) person and item parameters are estimated from a suitable IRT model, such as the model in Equation 9, (2) the model-implied probability for giving a correct response to each item in the item pair is derived from the model parameters, (3) the item residuals are calculated as the difference between the model-implied probability of a correct response and the observed response for each person, and (4) Q_3 is calculated as the Pearson correlation of the residuals of both items across persons. Yen (1993) noted that Q_3 is negatively biased given LI and suggested that a bias correction should be applied by subtracting the expected value of Q_3 given LI, which is $\frac{-1}{I-1}$, from all Q_3. The approach then contrasts the mean residual correlations between item pairs *kk'* referring to the same event with the mean residual correlations between item pairs *ll'* referring to different events to calculate the dependency measure D_{Q_3}:

$$D_{\mathrm{Q}_3} = \frac{1}{K} \sum_{k>k'} Q_3^{\mathrm{kk'}} - \frac{1}{L} \sum_{l>l'} Q_3^{\mathrm{ll'}}, \tag{10}$$

where K is the number of item pairs belonging to the same event and L is the number of item pairs belonging to different events. Given binding of event elements and a resulting stochastic dependency of the retrieval of event elements, within-event item residual correlations deviate from zero, whereas between-event item residual correlations are close to zero. Quantifying the dependency measure relative to the mean between-event item residual correlations allows for corrections of spurious item residual correlations that may be present in the data or be induced due to model misspecification, since these would affect both within- and between-event item residual correlations. The interpretation of the measure is equivalent to the one of the contingency-based measures, with zero indicating independence, positive values indicating positive dependency, and negative values indicating negative dependency.

Due to the sampling distribution of Q_3 being unknown (Chen & Thissen, 1997), the sampling distribution of D_{Q_3} is also unknown. In addition, the approach returns an overall measure of dependency for a given experimental condition or group. Thus, classical testing approaches (e.g., *t*-tests) are not applicable. For testing whether dependency or differences in dependency significantly differ from zero one can instead use a bootstrap approach. We use parametric bootstrapping, thus using the parameters estimated from the IRT model to generate data under the assumption that the data-generating model is true. When sampling from a unidimensional IRT model such as the one in Equation 9, this implies that the LI assumption holds and there is thus no dependency. By calculating the dependency measure or differences in dependency measures for each bootstrap sample one can generate distributions of the respective indices under the null hypothesis, from which *p* values for the observed indices can be derived. For testing for differences in dependency, the null hypothesis is that dependency in the compared conditions or groups is equal, but not necessarily zero. Thus, for this kind of test, one needs to generate data from a model that allows for dependencies in item responses. This can be achieved by sampling from a bifactor IRT model (see Gibbons & Hedeker, 1992; Wainer & Wang, 2000), which extends the model in Equation 8 by including additional, event-specific, latent traits λ, thus making the model multidimensional:

$$P(u_{\mathrm{ij}} = 1) = \gamma_{\mathrm{j}} + (1 - \gamma_{\mathrm{j}}) \frac{e^{\alpha_{\mathrm{j}}(\theta_{\mathrm{i}} - \beta_{\mathrm{j}}) - \alpha_{\mathrm{t(j)}}\lambda_{\mathrm{it(j)}}}}{1 + e^{\alpha_{\mathrm{j}}(\theta_{\mathrm{i}} - \beta_{\mathrm{j}}) - \alpha_{\mathrm{t(j)}}\lambda_{\mathrm{it(j)}}}}, \tag{11}$$

where λ is the event-specific latent trait of person *i* for event *t(j)* to which item *j* belongs. In accordance with the simplifications made to the model in Equation 9, this model can also be simplified:

$$P(u_{\mathrm{ij}} = 1) = g + (1 - g) \frac{e^{\theta_{\mathrm{i}} - \beta_{\mathrm{j}} - \lambda_{\mathrm{it(j)}}}}{1 + e^{\theta_{\mathrm{i}} - \beta_{\mathrm{j}} - \lambda_{\mathrm{it(j)}}}} \tag{12}$$

All latent traits in this model are mutually independent, and thus there is no conditional dependency in item responses referring to different events. The event-specific latent traits induce stochastic dependencies of item responses referring to the same event via their variance, with higher variances indicating higher dependencies (i.e., stronger event-specific effects). For informing the parametric bootstrap, one also needs to fit this model to the empirical data. Since experiments usually include several events, the model may quickly become highly dimensional, because an additional event-specific trait is required for each event. This is especially problematic considering the relatively small sample sizes typical of experiments compared to, for example, large-scale educational assessments. To mitigate this problem, it is advisable to reduce the number of parameters to be estimated by setting equality constraints on event-specific trait variances within experimental conditions or groups, also considering that events are usually randomly generated for each participant.

2.2.2 Nonparametric Variant

While the previously described IRT-based approach is parametric, Debelak and Koller (2020) proposed a nonparametric estimation procedure for Q_3, with which a nonparametric variant of the dependency measure ($D^{\mathrm{np}}_{\mathrm{Q}_3}$) can be calculated. The estimation procedure builds on the nonparametric testing framework by Ponocny (2001) and on a property of the Rasch model (Rasch, 1960) that marginal person and item sums are sufficient statistics for person and item parameters. Thus, the procedure involves the generation of bootstrap samples of artificial response matrices with the same marginal sums as the observed response matrix, using a Markov-Chain Monte-Carlo algorithm by Verhelst (2008). Then, the probability for a person giving a correct response to an item is calculated by averaging the respective responses in the generated artificial response matrices (u_{ij}) across bootstrap samples. Subsequently, nonparametric Q_3 statistics and $D^{\mathrm{np}}_{\mathrm{Q}_3}$ are calculated like their parametric counterparts (see Equation 10). In addition, one can calculate $D^{\mathrm{np}}_{\mathrm{Q}_3}$ for each bootstrap sample to derive p values for $D^{\mathrm{np}}_{\mathrm{Q}_3}$ and differences in $D^{\mathrm{np}}_{\mathrm{Q}_3}$.

2.3 Evaluation of Measurement Approaches

Given the availability of several approaches for measuring binding effects in episodic memory, the question remains which of these approaches are best suited for this task. While I already discussed some advantages of the IRT-based approach compared to the contingency-based approaches, such as the susceptibility to Simpson's paradox of the latter, in the first article (Schreiner & Meiser, 2022), we further evaluated the different approaches in terms of three important measurement properties (cf. J. Cohen, 1988): (1) bias of the estimates yielded by the different approaches, (2) Type I error rates, and (3) power. This was done for both tests against independence of individual estimates and tests for differences between experimental conditions. In addition, we investigated how susceptible the measurement properties of the different approaches are to variations in overall memory performance of the sample. Ideally, dependency measures should not be susceptible to memory performance to allow for dissociating dependency of the retrieval of event elements due to binding effects from higher memory performance.

For the evaluation, we conducted a Monte Carlo simulation in which we simulated an experimental within-subjects design with two experimental conditions, a total of $T = 30$ events (15 per condition), and five response options in the memory test (resulting in a stochastic guessing probability of $g = 0.2$). We simulated six test trials per event, which corresponds to testing each association in an event consisting of three elements in both directions (e.g., testing vendor–bread, bread–vendor, vendor–bakery, bakery–vendor, bread–bakery, and bakery–bread). Item responses were drawn from a bifactor IRT model (Gibbons & Hedeker, 1992; Wainer & Wang, 2000, see Equation 11), which allowed us to induce dependencies between item responses by adding event-specific latent traits (λ). We varied the sample size ($N = \{25, 50, 75, 100\}$), the dependency of the retrieval of event elements by varying the variances of the event-specific latent traits ($\text{Dep.} = \{0, 0.5, 1\}$)[2], differences in dependency between experimental conditions by increasing event-specific trait variances in the second condition relative to the first condition ($\text{Dep.}_{\text{diff}} = \{0, 0.5, 1\}$), and the overall level of memory performance in the sample by changing the mean of the general latent trait θ ($P = \{\text{-}2, 0, 2\}$, resulting in a proportion of 40%-42%, 59%-60%, and 75%-80% correct responses, respectively). This resulted in 108 simulation conditions, for each of which we conducted 1,000 replications. For the contingency-based approaches (D_{HB}, Q, and Q_{a}), we conducted one-sample t-tests against zero for testing against independence and paired t-tests for testing for differences between conditions. For the IRT-based approaches we used bootstrapping (parametric for D_{Q_3}, nonparametric for $D^{\text{np}}_{\text{Q}_3}$) for determining statistical significance, using 1,000 bootstrap samples, respectively (cf. Davison & Hinkley, 1997). All tests were one-tailed, since no negative dependencies can be induced using the bifactor model.

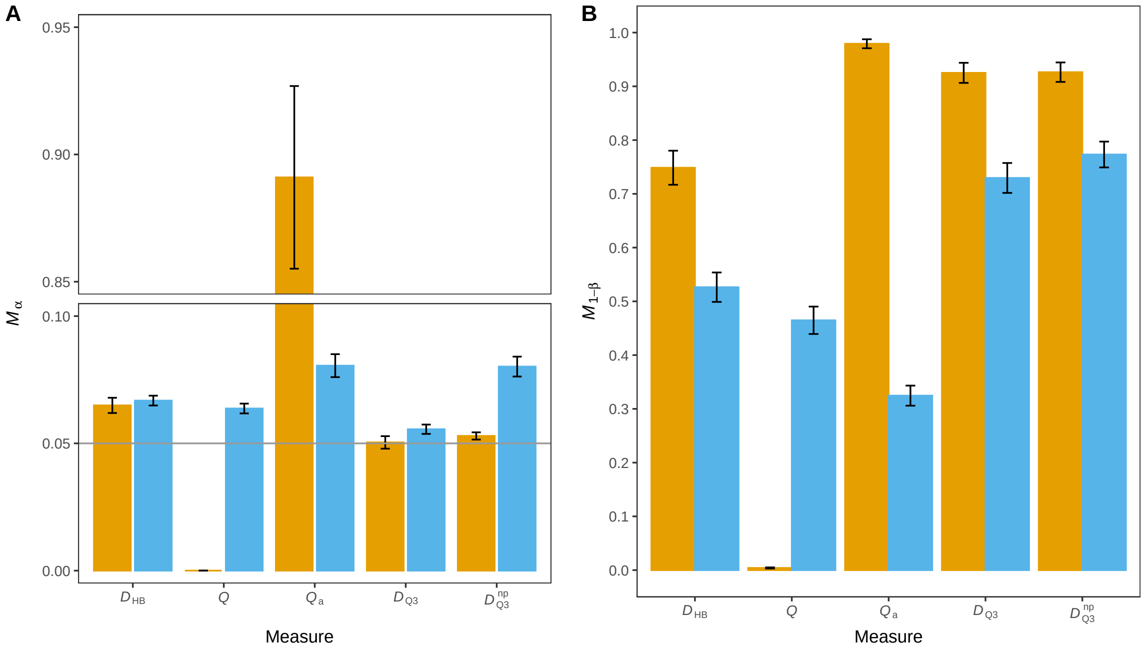
Figure 2 shows average Type I error rates and power of the different measures across simulation conditions. More detailed results are presented in the first article (Schreiner & Meiser, 2022). The simulation revealed that Q is negatively biased, whereas Q_{a} is positively biased. For both measures the bias increases with increasing memory performance. For Q_{a} this results in severely inflated Type I error rates that further increase with increasing memory performance. For Q this results in virtually no sensitivity to dependency when testing against independence (i.e., very low power).

---

[2] Note that for event-specific trait variances of zero the bifactor model reduces to a unidimensional model with the LI assumption (i.e., independence of item responses given the general latent trait).

Figure 2

Average Type I error rates (A) and power (B) of the different measures across simulation conditions when testing against independence and when testing for differences between conditions in the simulation study by Schreiner and Meiser (2022).



Note. For Type I error rates (A), the horizontal grey line indicates the nominal significance level. For tests against independence, values are also averaged across the simulated experimental conditions. Error bars represent ± *SEM*.

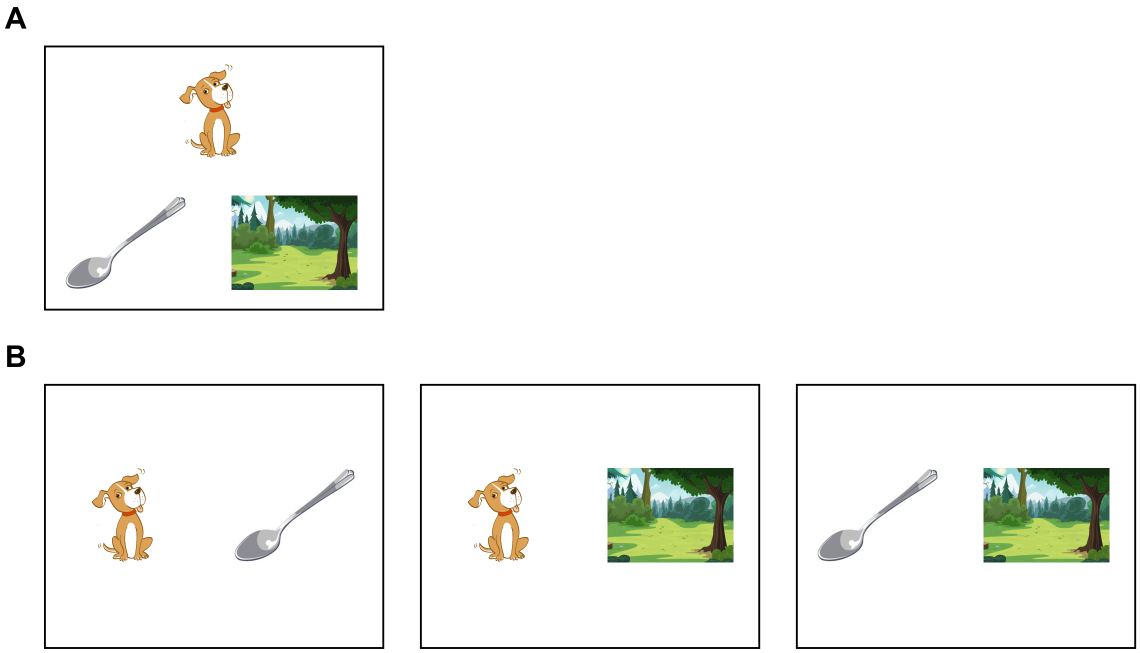
However, estimates of dependency differences between conditions are unbiased, but Q_{a} tends to yield slightly increased Type I error rates when testing for dependency differences. In terms of power, both Q and Q_{a} are inferior to the other approaches. D_{HB}, D_{Q_3}, and $D^{\mathrm{np}}_{\mathrm{Q}_3}$ are unbiased and not susceptible to memory performance given independence of item responses. All three measures show acceptable to good maintenance of the nominal significance level. Given dependency, the measures are affected by memory performance (D_{HB} less so than the IRT-based approaches), but this kind of susceptibility is less concerning, since it only occurs if there is a true effect, and thus only affects power, but not Type I error rates. The IRT-based approaches however yield higher power for detecting dependency than does D_{HB} and power is affected by memory performance to a similar degree for all three measures. When considering dependency differences and testing for differences between conditions, these results were largely mirrored. However, $D^{\text{np}}_{\text{Q}_3}$ yields increasing Type I error rates as dependency in the data increases. This is not the case for the other measures. In addition, given true differences in dependency between conditions, estimates of dependency differences from all measures shift closer to zero as dependency in the data increases, resulting in decreasing power with increasing dependency in the data.

As a complement to the simulation study, we reanalyzed an empirical dataset by James et al. (2020) using the different approaches to compare resulting inferences drawn from empirical data (using two-tailed testing). In their first experiment ($N = 45$), James et al. (2020) presented participants with 30 events, each consisting of 3 elements (an animal, an object, and a location) that were shown as cartoon illustrations and additionally named via audio recordings projected through headphones. Using a cued recognition test, there were six test trials per event (i.e., each association was tested in both directions). The experiment encompassed two within-subjects conditions: In a simultaneous encoding condition, all event elements were shown simultaneously, during a single learning trial (cf. Horner & Burgess, 2013). In a separated encoding condition (cf. Horner et al., 2015; Horner & Burgess, 2014), event elements were shown sequentially pairwise, across three temporarily divided learning trials (see Figure 3).

Two previous studies found a significant positive dependency of the retrieval of event elements in both the simultaneous and separated encoding condition that did not significantly differ between the conditions (Bisby et al., 2018; Horner & Burgess, 2014), suggesting binding effects of similar magnitude in both conditions. However, James et al. (2020) found a significant dependency only in the simultaneous encoding condition, but not in the separated encoding condition, with a significant difference between the conditions, suggesting binding effects were only present in the simultaneous encoding condition. All three studies employed the approach by Horner and Burgess (2013). Using the IRT-based approaches, both the parametric and nonparametric variant yielded a significant dependency in both conditions, but the dependency in the simultaneous encoding condition was significantly larger. These results are thus more consistent with the results by Bisby et al. (2018) and Horner and Burgess (2014) than are the results by James et al. (2020). The significant dependency in the separated encoding condition may be explained by the higher power for detecting dependencies of the IRT-based approaches compared to the approach by Horner and Burgess (2013). Q and Q_{a} yielded diverging results, with Q yielding no

Figure 3

Exemplary learning trials of a simultaneous encoding condition (A) and a separated encoding condition (B) for an event encompassing the elements dog, spoon, and forest.



Note. Pictures were taken from James et al. (2020), available at https://osf.io/vqzh8/. Learning trials in a separated encoding condition usually do not directly follow each other but are presented interleaved with learning trials from other events. This was also the case in James et al. (2020).

significant dependency in the simultaneous encoding condition and a significant negative dependency in the separated encoding condition, with a significant difference between the conditions, and Q_{a} yielding a significant dependency in both conditions, with no significant difference between the conditions. These divergent findings may be explained by the measures being biased.

In sum, the parametric IRT-based approach (D_{Q_3}) seems best suited for measuring the stochastic dependency of the retrieval of event elements as an indicator of binding effects, yielding unbiased estimates of dependency and dependency differences, good maintenance of Type I error rates, high power, and empirical inferences in accordance with previous findings. The nonparametric variant ($D^{\mathrm{np}}_{\mathrm{Q}_3}$) also generally performs well, but is prone to increased Type I error rates when testing for differences in dependency. Q and Q_{a} on the other hand seem unsuited for measuring binding effects, given their bias and associated problematic Type I error rates and power, and the strongly diverging empirical inferences drawn when using these measures. The approach by Horner and Burgess (2013) (D_{HB}) also seems to yield a suitable measure, albeit performing worse than D_{Q_3}. However, it may be particularly useful when person-specific estimates are required, for example when one wants to investigate the influence of continuous covariates such as age on the binding of event elements. Person-specific estimates are not yet provided by D_{Q_3}.

Besides the discussed advantages of the IRT-based approach and its insusceptibility to Simpson's paradox, the approach provides some additional advantages over the contingency-based approaches. While the latter are in essence descriptive, the IRT-based approach utilizes established and plausible modeling of meaningful psychological variables (e.g., memory performance as a latent trait and event-specific effects defined in terms of item residual correlations and as additional latent traits in bifactor models). Resulting person and item parameters can also be used for additional analyses and goals, for example for investigating participants' memory performance or the compilation of study materials if one wants to use the same events across participants, to ensure comparable difficulty of different events. The IRT-based approach, resting on individual item responses, can further be applied to a greater variety of test formats, such as free recall, in which the lack of cue-target pairs would render dependency pairs used in the contingency-based approaches arbitrary. Finally, the IRT-based approach can be extended to polytomous item responses, for example by using the rating scale (Andrich, 1978) or partial credit model (Masters, 1982) and then calculating item residual correlations from these models. A potential application with polytomous item responses may be the investigation of dependencies in confidence judgments in memory tasks.

3 Binding Structures

In Chapter 2, I identified suitable approaches, particularly an approach based on IRT, for measuring binding effects. In the following chapters, I will use this measure (D_{Q_3}) to investigate substantive research questions regarding the binding of event elements in episodic memory. An important question is the structure in which different event elements are bound together, since this is a fundamental aspect of information storage and retrieval in episodic memory. There exist different competing accounts of how bound event elements are structured.

3.1 Integrated Binding Accounts

Integrated binding accounts suggest that event elements are bound into a unitary representation, which can be accessed in a holistic manner. Tulving (1983) suggested that information regarding different event elements is stored in event engrams, which are discrete bound event representations. The hippocampus may act as a convergence zone that binds event elements into discrete engrams that can be retrieved by partial activation of event elements via pattern completion (Damasio, 1989; Marr, 1971; Moll & Miikkulainen, 1997). Furthermore, the integrative encoding hypothesis suggests that the hippocampus integrates newly encountered associations into existing overlapping ones, which ultimately leads to integrated representations containing all event elements (Shohamy & Wagner, 2008; Zeithamova et al., 2012). As a consequence of integrated binding structures, one can assume that there are no longer individual associations between individual event elements. Instead, event elements are fully integrated into a superordinate memory structure that can only be accessed holistically. Therefore, asymmetries in binding strength are not possible.

3.2 Pairwise and Hierarchical Binding Accounts

Other accounts suggest that event elements may be bound together in a network of pairwise associations that potentially allows for asymmetries in binding strength. Relational memory theory suggests that the hippocampus flexibly binds elements into a network-like structure depending on task demands (N. J. Cohen & Eichenbaum, 1993; Eichenbaum, 1999; see also Eichenbaum & Cohen, 1988, 2001). With ensemble encoding, associations may be stored as overlapping neural ensembles, but these ensembles may remain distinct rather than forming a unitary representation

(Cai et al., 2016). The theory of event coding (TEC) suggests that feature codes, being codes of stimuli, are perceptually activated and bound into event files (Hommel, 1998, 2009; see also the recently proposed binding and retrieval in action control [BRAC] framework; Frings et al., 2020). Event files consist of multiple local interconnections of feature codes (Hommel, 1998, 2004) and connections may be sparse, since not all possible connections are necessarily formed (Moeller et al., 2019). Feature codes may also contribute to the event file with varying degrees (Hommel et al., 2001). Thus, binding asymmetries are possible in the TEC. The Span-Cospan model of episodic memory (Healy & Caudell, 2019) suggests that event elements are bound into higher-order representations of event segments, which may be bound into further higher-order representations, up to a representation encompassing the entire event. However, holistic access to individual event segments is maintained. Representations and connections can vary in strength, and thus the model also allows for asymmetries in binding strength. These accounts suggest that event elements may be bound into a hierarchical binding structure, in which some elements are preferentially bound to other elements. Such structures may be enabled by systematic variations in binding strength (including binding strengths of zero, i.e. no direct bindings being formed between some event elements).

3.3 Testing an Integrated Against a Hierarchical Binding Account

Schreiner, M. R., Meiser, T., & Bröder, A. (2022). The binding structure of event elements in episodic memory and the role of animacy. *Quarterly Journal of Experimental Psychology.* Advance online publication. https://doi.org/10.1177/17470218221096148

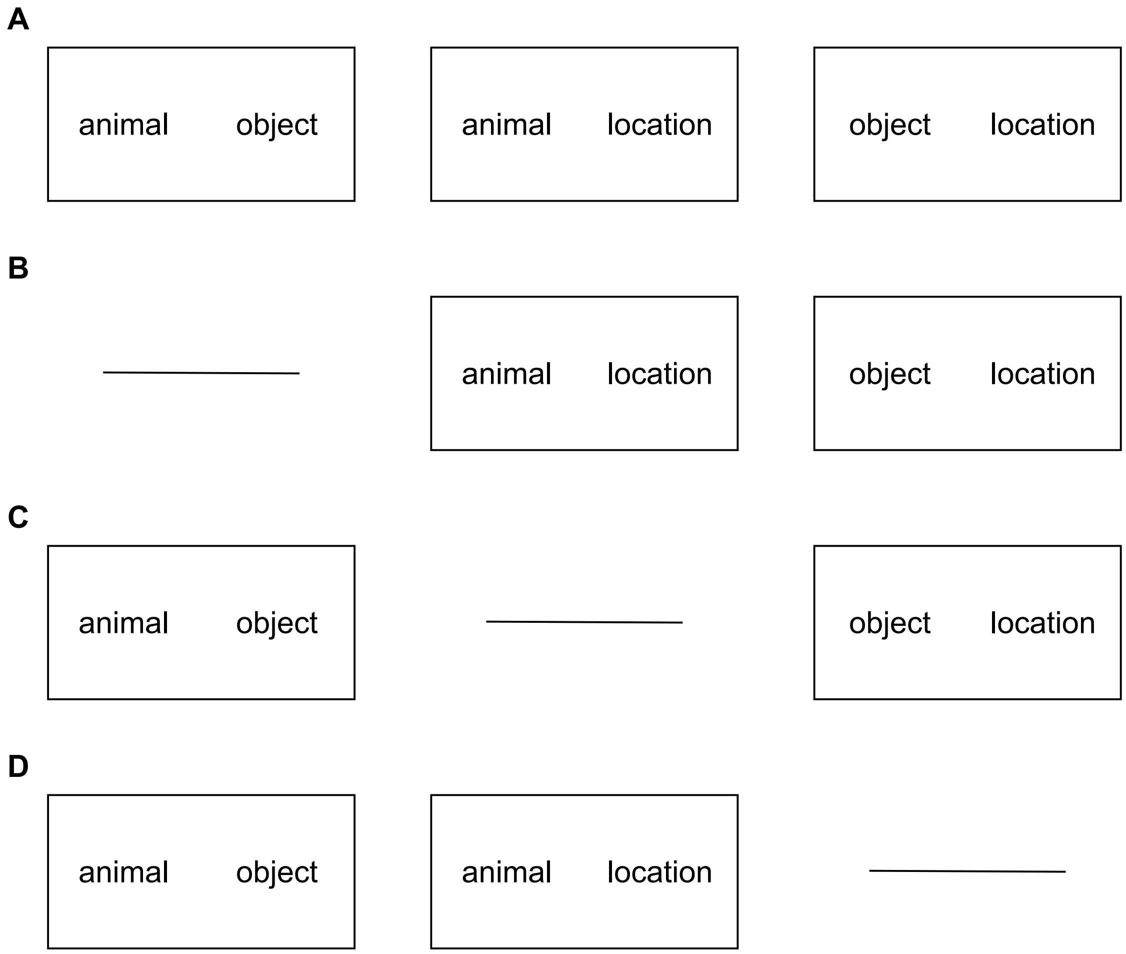
In the second article (Schreiner, Meiser, & Bröder, 2022), we empirically tested an integrated against a hierarchical binding structure using an experimental paradigm suitable to distinguish between these different binding structures in three experiments. Participants were presented several events consisting of three elements presented as nouns and were instructed to imagine these elements as part of a scence and interacting in a meaningful manner. In Experiment 1, all events consisted of an animal, an object, and a location (animacy condition). Experiments 2 and 3 additionally included a non-animacy condition in which events consisted of two types of objects (means of transportation and tools) and a location. Animacy condition was a within-subjects factor in Experiment 2 and a between-subjects factor in Experiment 3. We manipulated animacy because we suspected it to facilitate the formation of hierarchical binding structures, but the rationale behind different animacy conditions is primarily discussed in Chapter 4. We employed the separated encoding paradigm (Horner et al., 2015; Horner & Burgess, 2014), in which event elements are presented sequentially pairwise. In addition to a closed-loop (CL) condition (coherent encoding episodes), in which all possible pairwise associations are presented (see Figure 4A, see also Figure 3B), we additionally included three open-loop (OL) conditions, in which we excluded one of the possible pairwise associations from presentation, respectively (non-coherent encoding episodes, see Figure 4B-D). In a subsequent test phase, participants were presented with an event element as a cue and had to select the associated target element that belonged to the same event as the cue from six response alternatives.

While we included the closed-loop condition to replicate previous findings showing that event elements can be bound across several temporarily divided encoding episodes (Bisby et al., 2018; Horner et al., 2015; Horner & Burgess, 2014; Joensen et al., 2020), integrated and hierarchical binding structures make different predictions regarding the pattern of stochastic dependencies of the retrieval of event elements across the different open-loop conditions. Because an integrated binding structure consists of a unitary event representation that can only be accessed holistically, dependency should not vary across the open-loop conditions, since all associations, including the one not presented, are fully integrated and should be readily retrieved with all the other associations (or integration may fail for non-coherent encoding episodes; cf. Horner et al., 2015; Horner & Burgess, 2014). However, in a hierarchical binding structure there are systematic variations in binding strength. Thus, associations between more strongly bound event elements should more strongly contribute to a stochastic dependency of the retrieval of event elements than associations between less strongly bound event elements. Excluding associations from presentation that would more strongly contribute to dependency should therefore diminish dependency compared to excluding associations that would less strongly contribute to dependency, and thus dependency should vary across the different open-loop conditions.

The dependency results for the three experiments are shown in Figure 5. While we replicated a significant positive dependency in two out of the three experiments (in Experiments 1 and 3), suggesting that event elements can indeed be bound across several temporarily divided encoding episodes, at least if encoding episodes are coher-

Figure 4

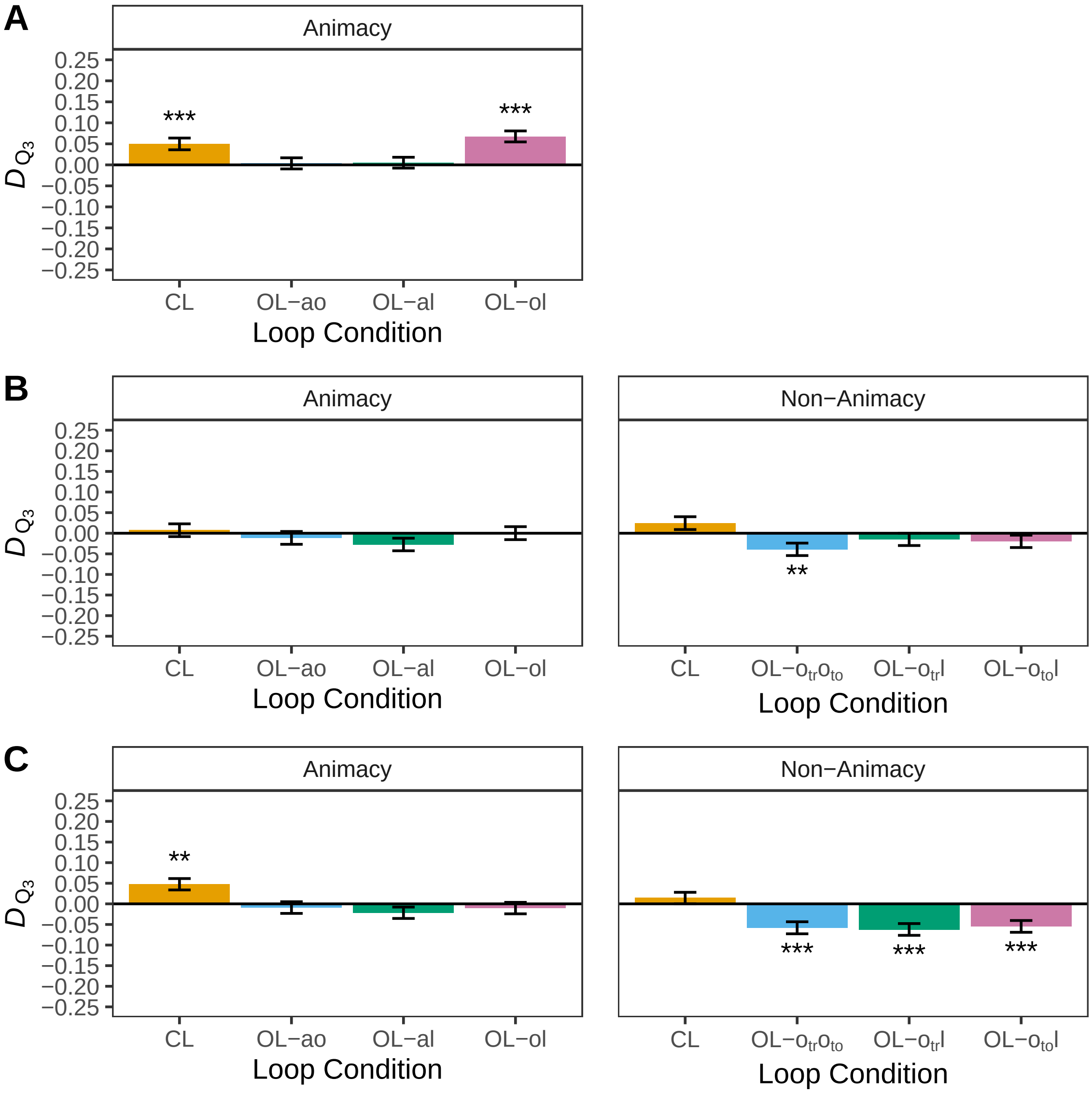
Learning trials for the closed-loop condition (A) and for different open-loop conditions (B-D) for an event consisting of an animal, an object, and a location in the separated encoding paradigm.



Note. Horizontal lines indicate that the respective learning trial did not occur in the respective condition. Learning trials for an event did not directly follow each other but were presented interleaved with learning trials from other events.

Figure 5

Dependency of the retrieval of event elements by animacy and loop condition for Experiment 1 (A), Experiment 2 (B), and Experiment 3 (C) of Schreiner, Meiser, and Bröder (2022).



Note. $^{**}p < .01$, $^{***}p < .001$. CL = closed-loop, OL-ao = open-loop with association animal – object excluded, OL-al = open-loop with association animal – location excluded, OL-ol = open-loop with association object – location excluded, OL-$o_{tr}o_{to}$ = open-loop with association means of transportation – tool excluded, OL-o_{tr}l = open-loop with association means of transportation – location excluded, OL-o_{to}l = open-loop with association tool – location excluded. Error bars represent ± *SE*.

ent, results regarding the binding structure were mixed. In Experiment 1 there was a significant positive dependency in the open-loop condition in which the association object – location was excluded that did not significantly differ from the dependency in the closed-loop condition and was significantly larger than the dependency in the other open-loop conditions in which the animal – object or animal – location association was excluded. This result pattern suggests that event elements are bound in a hierarchical manner (cf. Cai et al., 2016; N. J. Cohen & Eichenbaum, 1993; Eichenbaum, 1999; Healy & Caudell, 2019; Hommel et al., 2001), with event elements being preferentially bound to the animal element. In the animacy condition of Experiment 3 however, we did not find a significant dependency in any of the open-loop conditions. This finding is more consistent with an integrated binding structure (Damasio, 1989; cf. Horner et al., 2015; Horner & Burgess, 2014; Joensen et al., 2020; Marr, 1971; Moll & Miikkulainen, 1997; Shohamy & Wagner, 2008; Tulving, 1983; Zeithamova et al., 2012), with integration failing for non-coherent encoding episodes (cf. Horner et al., 2015; Horner & Burgess, 2014).

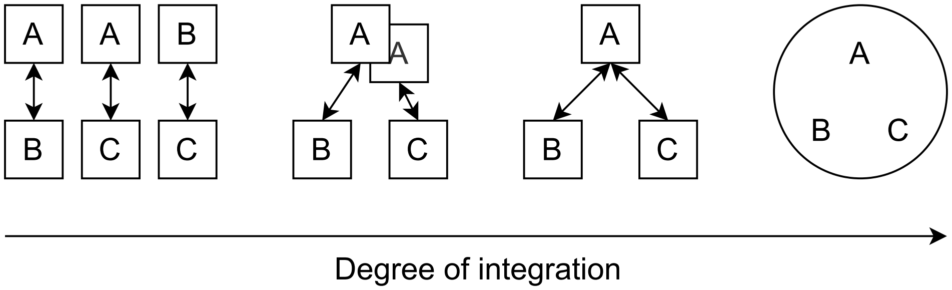
Experiment 2 was not particularly informative regarding a distinction between an integrated and a hierarchical binding structure, since we did not find a significant dependency in almost all conditions, not even in the closed-loop conditions. This suggests that in this experiment participants may have formed independent pairwise representations of event elements that were not integrated into a superordinate memory representation. This may have been the case because, in Experiment 2, events could vary in the composition of their elements, since they consisted of an animal, an object, and a location in the animacy condition, and two types of objects, and a location in the non-animacy condition. Since animacy condition was manipulated within-subjects, participants may have been less aware of the underlying event structure compared to the other experiments, which may have interferred with the formation of abstract representations of event structures. These may, however, be beneficial for binding (Morton et al., 2020; see also Kumaran, 2013; Kumaran & Ludwig, 2013).

Interestingly, we found significant negative dependencies in the open-loop conditions of the non-animacy condition in Experiment 3 (and also in one condition in Experiment 2). This may suggest that pairs of event elements were encoded as distinct overlapping events and representations were then driven apart by pattern separation processes in the hippocampus (see Zotow et al., 2020). An alternative explanation may be that the selective retrieval of one event element may have inhibited the retrieval of other event elements at the time of testing (cf. Horner & Burgess, 2013) and negative dependencies may thus be due to retrieval-induced forgetting (Anderson et al., 1994).

In sum, the results of the three experiments yielded unclear evidence for specific binding structures. This may suggest that different types of binding structures can be formed (and may even exist in parallel). For example, it is conceivable that there may exist a continuum of increasingly integrated memory representations, as depicted in Figure 6. Weakly integrated representations may be represented by independent pairwise bindings and may have occurred in Experiment 2. (Initially) overlapping pairwise representations may be slightly more integrated and may have occurred in the non-animacy condition of Experiment 3. Further up the continuum may then be hierarchical binding structures, as suggested by the results of Experiment 1, and ultimately fully integrated binding structures, as suggested by the results in the animacy condition of Experiment 3. While more integrated representations may intuitively be beneficial, they may also come with costs and may lead to seemingly paradoxical effects. For example, while initially overlapping pairwise representations may be considered higher up the integration continuum than independent pairwise representations, they may lead to negative dependencies due to pattern separation processes driving representations apart. This however, is an adaptive property of the hippocampus that reduces interference between similar representations in memory (Guzowski et al., 2004; Neunuebel & Knierim, 2014; Yassa & Stark, 2011). Also, while integrated representations may require less storage space than hierarchical representations, because all event elements are bound into a unitary representation, individual associations are no longer accessible in integrated representations. Thus, if a memory trace is too weak or fades over time, accessibility to the whole representation may be lost (see e.g., Joensen et al., 2020), whereas for hierarchical representations some associations may still be accessible and help in inferring the remaining associations. The degree of integration a memory representation achieves may be influenced by several moderators.

Figure 6

Schematic depiction of representations for an event consisting of three event elements (A, B, and C) along an integration continuum. From left to right, the depicted representations are independent pairwise, (initially) overlapping pairwise, hierarchical, and integrated representations.



4 Moderators of Binding Processes

In Chapter 3, I presented and discussed evidence suggesting that the structure in which event elements are bound together may vary, such that different binding structures with different degrees of integration are possible. From that the question follows what causes some events to achieve higher degrees of integration than others. In this chapter, I will thus investigate moderators of the binding of event elements.

So far, such moderators have only been scarcely investigated. James et al. (2020) found that, when presenting event elements as words, the addition of spoken words to the presentation of written words (i.e., multimodal instead of unimodal presentation) disrupted binding. Also, the use of picture stimuli disrupted binding compared to written stimuli, suggesting an effect of the modality of the presentation of event elements. Further, there is some evidence that knowledge or awareness regarding the structure of an event, such as the number and types of elements making up an event, facilitates binding (Kumaran, 2013; Kumaran & Ludwig, 2013; Morton et al., 2020). Our findings in Schreiner, Meiser, and Bröder (2022), in which we found no binding effects in an experiment in which events could consist of different sets of element types (varying event composition) compared to experiments in which they always consisted of the same set of element types (fixed event composition) corroborates this evidence. Awareness regarding the structure of an event may enable people to map representations on a latent geometric space, facilitating integration and enabling vector-based retrieval and inference (Morton et al., 2020).

4.1 Animacy

The presence of an animate element in an event may be another moderator facilitating the formation of coherent memory representations. Animate entities are living things that are capable of independent movement and able to change direction without warning (Bonin et al., 2015). Animacy is an important factor influencing human cognition (Nairne et al., 2013, 2017) that may be explained by selective pressure shaping our ancestors' memory system (Nairne et al., 2007, 2008). In this context, animacy may be an important survival-related factor. For example, animate entities may be potential prey, predators, or sexual partners (Bonin et al., 2015; Nairne et al., 2017). Regarding memory performance, an animacy effect has been commonly found, such that words with an animate referent are remembered better than words with an inanimate referent (e.g., Li et al., 2016; Nairne et al., 2013; VanArsdall et al., 2015).

This effect has been robustly found across a variety of test formats, including free recall (Bonin et al., 2015; Leding, 2019; Li et al., 2016; Madan, 2021; Nairne et al., 2013; Popp & Serra, 2016), recognition (Bonin et al., 2014; Leding, 2020; VanArsdall et al., 2013), and judgments of learning (DeYoung & Serra, 2021; Li et al., 2016). In cued recall tests, evidence for an animacy effect has been mixed, with some studies finding an effect (DeYoung & Serra, 2021; Laurino & Kaczer, 2019; VanArsdall et al., 2015) and some studies even finding an opposite effect (Kazanas et al., 2020; Popp & Serra, 2016). These inconsistent findings may be explained by variability in within-pair similarity of the word pairs learned in cued recall tasks (Serra & DeYoung, 2022).

Beyond enhancing memory performance, in the second article (Schreiner, Meiser, & Bröder, 2022), we found evidence that animacy also facilitates the binding of event elements in episodic memory. In Experiment 1, in which events consisted of an animal, an object, and a location, we found a significant positive dependency of the retrieval of event elements when all possible pairwise associations between event elements were presented (closed-loop condition) and when an association not involving the animate element was excluded from presentation (one of the open-loop conditions), whereas we did not find dependencies when an association involving the animate element was excluded (see Figure 5A). This suggests a hierarchical binding structure, in which the inanimate event elements (object and location) are preferentially bound to the animate element (the animal).

In the subsequent experiments we manipulated animacy by constructing events that either included an animate element (as was the case in Experiment 1, animacy condition) or did not include an animate element (non-animacy condition). In the non-animacy condition, event elements consisted of two types of objects (a means of transportation and a tool) and a location. While manipulating animacy within-subjects yielded uninformative results, likely due to resulting varying event compositions and reduced awareness regarding the underlying event structure (cf. Kumaran, 2013; Kumaran & Ludwig, 2013; Morton et al., 2020, see Chapter 3), manipulating animacy between-subjects yielded strongly diverging result patterns between the animacy and non-animacy condition (see Figure 5C). While the result pattern in the animacy condition suggests an integrated binding structure (there was a significant positive dependency in the closed-loop condition, but no significant dependencies in the open-loop conditions), there were negative dependencies in the open-loop conditions of the non-animacy condition, potentially suggesting that overlapping pairwise representations were formed (cf. Zotow et al., 2020) and no signficant dependency in the closed-loop condition. These results suggest that representations with a higher degree of integration have been achieved in the animacy compared to the non-animacy condition.

In sum, the results suggest that the presence of an animate element in an event enables higher degrees of integration, and thus facilitates the formation of more coherent memory representations. The presence of an animate element may either provide a critical element in hierarchical binding structures, to which other event elements are preferentially bound (cf. Experiment 1), or facilitate the full integration of event elements into an integrated binding structure (cf. Experiment 3). In a supplemental analysis we found that facilitating effects of animacy only occurred for events for which an association involving the animate element was presented first. Animate elements may thus exert their influence by providing a potential agent in an event.

4.2 Agency

Schreiner, M. R., Bröder, A., & Meiser, T. (2022). *Agency effects on the binding of event elements in episodic memory.* Manuscript submitted for publication.

The facilitating effects of animacy on the binding of event elements in episodic memory may be due to animate elements providing a potential agent in an event, and thus effects may actually be driven by *agency.* Agency can be defined as "acting or having the capacity to act autonomously in a given environment" (Suitner & Maass, 2016, p. 248; see also Hitlin & Elder, 2007). While this definition certainly applies to animate entities, agency may also extend to inanimate entities (Johnson & Barrett, 2003; Lowder & Gordon, 2015). In that sense, agency can be considered to be a property of animate entities, but animacy may only be one of several factors contributing to agency. For example, the active performance of an action may be another factor contributing to an entity's perceived agency (e.g., a hunting fox may be ascribed more agency than a fox laying passively on the ground or a rabbit being hunted). Agency may thus be a more proximate explanation for the facilitating effects of animacy on the binding of event elements.

In the third article (Schreiner, Bröder, & Meiser, 2022), we tested whether agency, beyond animacy, facilitates the binding of event elements in five experiments. Participants were presented several events consisting of three event elements. Event elements were either three types of objects (a means of transportation, a tool, and a food item, Experiments 1 and 2) or three types of animals (a mammal, a bird, and an insect, Experiments 3-5) to avoid confounding with animacy and were presented as nouns. In Experiments 1-3 we employed the separated encoding paradigm (Horner et al., 2015; Horner & Burgess, 2014), which we also used in Schreiner, Meiser, and Bröder (2022), but this time we only included the closed-loop condition, in which all possible pairwise associations are presented[3]. In Experiments 4 and 5 we employed the simultaneous encoding paradigm (Horner & Burgess, 2013), in which all event elements are presented in a single learning trial. While the separated encoding paradigm provides a stricter test of binding, because event elements need to be bound across temporarily divided encoding episodes and binding effects are thus less likely confounded with covariations in perceptual variables, binding effects in the simultaneous encoding paradigm tend to be more robust (see James et al., 2020) and the simultaneous presentation of event elements is closer to how events are naturally experienced. Thus, in Experiments 1-3 there were two event elements per learning trial and three learning trials per event, and in Experiments 4 and 5 there were three event elements per learning trial and only one learning trial per event.

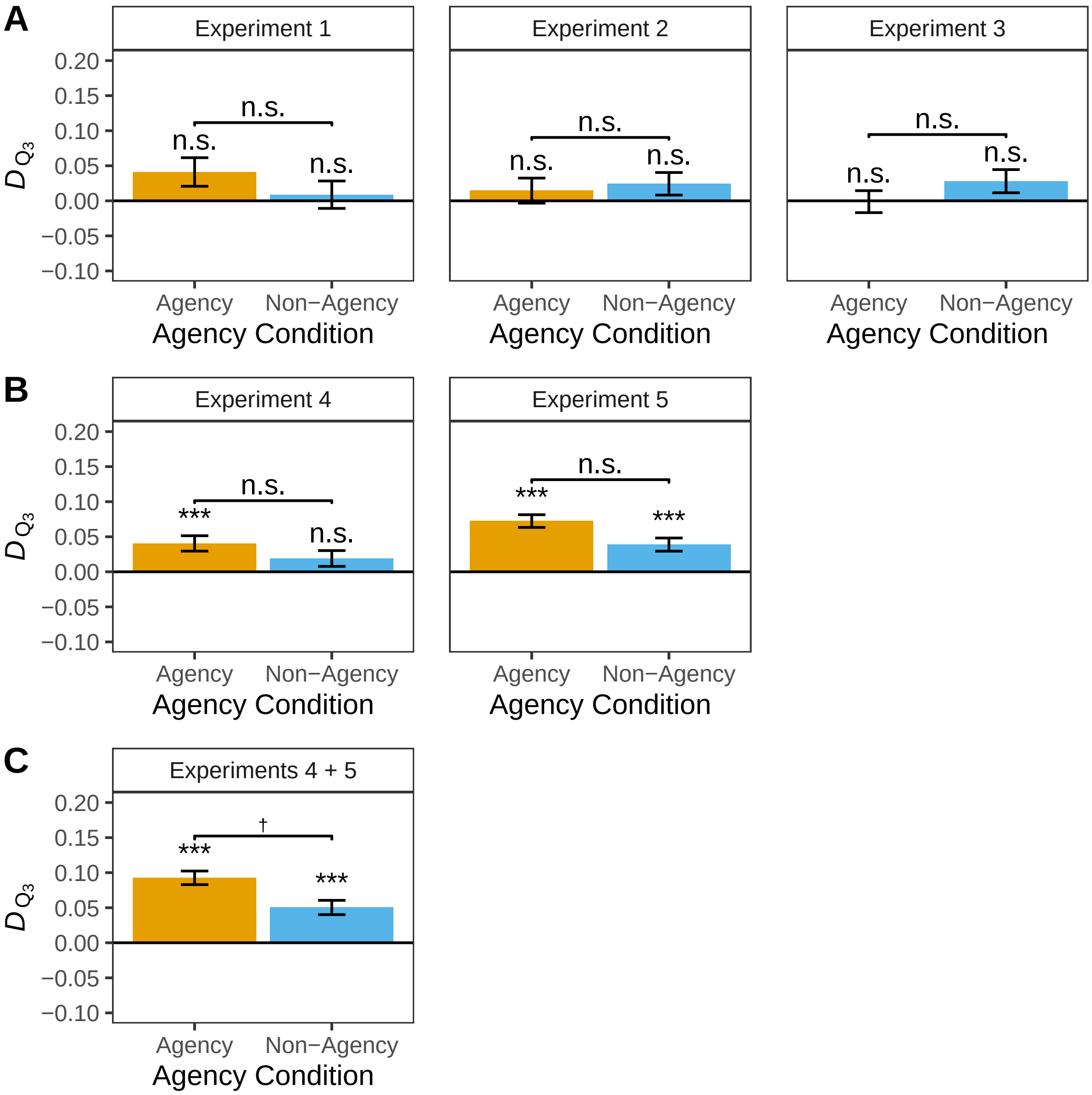
Event elements were presented embedded in sentences and we used a linguistic agency manipulation. In sentences with interpersonal action verbs (e.g., *hit*) the agent tends to be the grammatical subject, whereas the patient of the action tends to be the grammatical object (Kasof & Lee, 1993). In addition, the grammatical subject is perceived as more agent-like than the grammatical object (Kako, 2006) and the agent is given greater causal weight than the patient (Brown & Fish, 1983; Kassin & Lowe, 1979). We thus constructed sentences such that, in the agency condition, one of the event elements (the agent) served as the grammatical subject in a transitive active sentence (e.g., *The dog grabs the eagle.*), whereas the non-agentic element(s) served as the grammatical object(s). In the non-agency condition, we used passive sentences (e.g., *The dog and the eagle are being grabbed.*), in which the grammatical subject is not an agent (Kako, 2006). Such passive sentences were also used for sentences in the agency condition that did not include the agent element (this could only occur in the separated encoding paradigm). In a subsequent test phase, participants were presented with an event element as a cue and had to select the associated target element that belonged to the same event as the cue from six response alternatives, like in Schreiner, Meiser, and Bröder (2022).

---

[3] Experiment 2 also encompassed additional open-loop conditions to examine the binding structure of event elements, but yielded uninformative results concerning this question.

Figure 7

Dependency of the retrieval of event elements by agency condition for experiments employing the separated encoding paradigm (A), for experiments employing the simultaneous encoding paradigm (B) and for the aggregate analysis of Experiments 4 and 5 including only participants with above-median performance (C) of Schreiner, Bröder, and Meiser (2022).



Note. ***p < .001 (two-tailed), †p < .05 (one-tailed), n.s. = non-significant. Error bars represent ± *SE*. For Experiment 2 only data for the closed-loop conditions are shown.

The dependency results for the five experiments are shown in Figure 7. In experiments in which we employed the separated encoding paradigm (Experiments 1-3) we did not find any significant dependency of the retrieval of event elements in the agency and non-agency conditions (see Figure 7A), and thus no evidence for binding effects. This is at odds with results of previous studies, which found that binding effects also occur for (coherent) temporarily divided encoding episodes (Bisby et al., 2018; Horner et al., 2015; Horner & Burgess, 2014; Joensen et al., 2020; Schreiner, Meiser, & Bröder, 2022). The main difference between our experiments in Schreiner, Bröder, and Meiser (2022) and these previous studies is that we presented event elements embedded in sentences instead of presenting them as individual words or pictures. Presenting individual stimuli may allow participants to freely associate them, which may be further encouraged by the imagery instruction participants typically receive in these experiments. This may facilitate the binding of event elements compared to the more prestructured presentation of event elements embedded in sentences, which may inhibit participants' ability to freely associate them. Since our linguistic agency manipulation relied on the presentation of event elements embedded in sentences, it may not have not worked well in combination with the separated encoding paradigm.

In experiments in which we employed the simultaneous encoding paradigm (Experiments 4 and 5), we found significant positive dependencies of the retrieval of event elements (see Figure 7B), and thus evidence for binding effects. In Experiment 4 this was only the case in the agency condition. In Experiment 5, which had a larger sample size and a slightly longer presentation duration, this was the case in both the agency and non-agency condition. However, while the result pattern pointed in the expected direction (a higher dependency in the agency than in the non-agency condition), the difference in dependency between the two conditions was non-significant in both experiments.

It is noteworthy that memory performance in all experiments was quite poor. In Experiments 4 and 5, the average proportion of correct responses in the agency condition was 24%-26% and 23% in the non-agency condition. This is much lower than the memory performance in previous studies. For example, in our experiments in Schreiner, Meiser, and Bröder (2022) the average proportion of correct responses ranged from 38%-49% and in the experiments by Horner and Burgess (2013) it ranged from 57%-71%. As we showed in Schreiner and Meiser (2022), lower levels of memory performance are associated with lower power for detecting binding effects and differences in binding effects between conditions. It may thus have been the case that the low memory performance in the experiments prevented the reliable detection of a difference in dependency between the agency and non-agency condition.

Consequently, we performed a supplemental analysis in which we only included participants with above-median memory performance in their respective condition in Experiments 4 and 5 and aggregated the data from the two experiments to account for the reduced sample size. With this analysis, we found significant positive dependencies in both the agency and non-agency condition and, importantly, the difference in dependency between the conditions was also significant (see Figure 7C). The dependency was higher in the agency than in the non-agency condition. This finding corroborates the descriptive result patterns of Experiments 4 and 5 and suggests a facilitating effect of agency on the binding of event elements that may have been concealed due to low memory performance in the experiments.

In sum, the results of the experiments hinted at a facilitating effect of agency on the binding of event elements. While the results from the experiments with sequential pairwise event presentation were inconclusive, this may have been due to this presentation format being not particularly well suited for investigating the research question in combination with the linguistic agency manipulation used. Since event elements were embedded in sentences, this may have inhibited participants' ability to freely associate them. The results from the experiments with simultaneous event presentation yielded evidence for binding effects, but only descriptive evidence for a facilitating effect of agency. More concrete evidence only emerged in a supplemental analysis with aggregated data and only including participants with above-median memory performance. Thus, while the results are not very clear, they hint at a facilitating effect of agency and this effect may have been concealed due to low memory performance in the experiments. Therefore, agency may indeed be a more proximate explanation for the facilitating effects of animacy on the binding of event elements in Schreiner, Meiser, and Bröder (2022) and the presence of an agentic element in an event may facilitate the formation of more coherent memory representations.

5 Conclusion

The ability to bind together different elements of an event into a coherent memory representation is a fundamental property underlying episodic memory. Yet, little is known about how these binding processes work or how to properly measure them. In this thesis, I introduced and evaluated a novel approach for measuring binding processes in episodic memory using behavioral data. I then used this approach to examine the structure in which different event elements are bound together and moderators of binding processes.

Given that an event's constituent elements are successfully bound together, there should be a stochastic dependency of the retrieval of event elements (Arnold et al., 2019; Boywitt & Meiser, 2012a, 2012b; Horner et al., 2015; Horner & Burgess, 2013, 2014; Meiser & Bröder, 2002; Starns & Hicks, 2005, 2008). By modeling this dependency one can derive measures of binding effects. However, earlier modeling approaches (see Burton et al., 2019; Horner & Burgess, 2013, 2014; Yule, 1912) are contingency-based and come with a number of limitations, such as susceptibility to Simpson's paradox (Hintzman, 1972, 1980; Simpson, 1951). In the first article (Schreiner & Meiser, 2022; see also Schreiner, Meiser, & Bröder, 2022), we introduced a novel approach based on item response theory (Lord, 1980; Lord & Novick, 1968) that overcomes some limitations of previous approaches. For example, because the approach takes individual item responses instead of aggregate contingency tables as input, it is not susceptible to Simpson's paradox. In an evaluation, the approach yielded unbiased estimates, good maintenance of Type I error rates and high power for detecting binding effects, outperforming the contingency-based approaches. While dependency estimates scaled with memory performance, this only affected power but not Type I error rates. One drawback of the approach is that, while it takes person differences into account, it provides an overall measure of dependency, whereas the contingency-based approaches provide person-specific dependency estimates. In the subsequent articles we used this IRT-based approach to examine important properties of binding in episodic memory.

In the second article (Schreiner, Meiser, & Bröder, 2022), we investigated the structure in which different event elements are bound together. There are competing accounts of binding structures in the literature. One class of accounts suggests an integrated binding structure, in which all event elements are bound into one unitary representation or engram (Damasio, 1989; Marr, 1971; Moll & Miikkulainen, 1997;

Shohamy & Wagner, 2008; Tulving, 1983; Zeithamova et al., 2012). Another class of accounts suggest a system of pairwise bindings in which asymmetries in binding strength are possible, making possible a hierarchical binding structure, in which event elements are preferentially bound to particular types of elements (Cai et al., 2016; N. J. Cohen & Eichenbaum, 1993; Eichenbaum, 1999; Healy & Caudell, 2019; Hommel et al., 2001). We extended the separated encoding paradigm (Horner et al., 2015; Horner & Burgess, 2014), in which event elements are presented sequentially pairwise across several temporarily divided encoding episodes (either all possible associations are shown [coherent encoding episodes] or one association is excluded from presentation [non-coherent encoding episodes]), with several non-coherent encoding conditions, in which we consistently excluded one of the possible associations from presentation. In doing so, we replicated previous findings demonstrating that binding across temporarily divided encoding episodes is possible (Bisby et al., 2018; Horner et al., 2015; Horner & Burgess, 2014; Joensen et al., 2020). We further found evidence for different binding structures, including both integrated, hierarchical, and, potentially, overlapping pairwise representations. This suggests that different binding structures can be possibly formed and implies the need for an overarching account bridging the accounts in favor of integrated binding structures and those in favor of a system of pairwise bindings. For example, memory representations may lie on a continuum with varying degrees of integration. Different testing demands or contexts may elicit different structures, such that how events are represented may not be fixed, but vary dynamically based on the specific demands or contexts at play. This would be consistent with relational memory theory (N. J. Cohen & Eichenbaum, 1993; Eichenbaum, 1999). Thus, several moderators may influence the degree of integration and therefore the binding structure of a memory representation.

Also in the second article (Schreiner, Meiser, & Bröder, 2022), we investigated the influence of animacy on the binding of event elements. Animacy has previously only been investigated in relation to memory performance, but not regarding the coherence of memory representations. Regarding memory performance, an animacy effect has been robustly found, such that words describing an animate entity are remembered better than words describing an inanimate entity (e.g., Li et al., 2016; Nairne et al., 2013; VanArsdall et al., 2015). Extending these findings, we found evidence that the presence of an animate element in an event also facilitates binding, leading to more coherent memory representations than if an event is only comprised of inanimate elements.

In the third article (Schreiner, Bröder, & Meiser, 2022), we investigated the influence of agency on the binding of event elements. While robust effects were not seen, the pattern of results hinted at a facilitating effect of the presence of an agentic element in an event on binding. Agency effects may thus underlie the previously found facilitating effects of animacy. These may have occurred because the presence of an animate element in an event provides a potential agent. However, animacy may only be one of several factors contributing to agency (see e.g., Johnson & Barrett, 2003; Lowder & Gordon, 2015). In addition, our results suggest that the opportunity to freely associate event elements may facilitate binding. We only found binding effects when event elements were presented simultaneously, but not when they were presented sequentially pairwise, although binding effects have been found under such circumstances in previous studies (Bisby et al., 2018; Horner et al., 2015; Horner & Burgess, 2014; Joensen et al., 2020; Schreiner, Meiser, & Bröder, 2022). However, in Schreiner, Bröder, and Meiser (2022) we presented event elements embedded in sentences instead of presenting them as individual stimuli. This presentation format may have made the described scenes more prestructured and inhibited participants ability to freely associate event elements. This also corroborates evidence suggesting that additional processes may be required when binding event elements across temporarily divided encoding episodes compared to binding within a single encoding episode (see James et al., 2020).

Our findings on moderators of binding processes in episodic memory add to a limited number of previous findings. These suggest that multimodal compared to unimodal presentation of stimuli disrupts binding and that written stimuli faciliate binding compared to picture stimuli (James et al., 2020). The latter finding may also be consistent with our findings suggesting that free association faciliates binding (Schreiner, Bröder, & Meiser, 2022), since written stimuli may be more freely associated than picture stimuli. Further, awareness regarding the structure of an event, such as the number and types of its constituent elements, may facilitate binding (Kumaran, 2013; Kumaran & Ludwig, 2013; Morton et al., 2020). This is also corroborated by our findings in Schreiner, Bröder, and Meiser (2022). There, we only found binding effects with fixed event compositions, in which all events comprised the same set of element types, but not with varying event compositions, in which events could comprise different sets of element types. Participants' awareness regarding the underlying event structure may have been reduced in the case of varying compared to fixed event compositions.

To conclude, the present thesis provides a novel measure for investigating binding processes in episodic memory and insights into some properties of these binding processes. I demonstrated that the novel measure is well-suited to probe binding effects. Moreover, I provided evidence that event elements may be bound into different types of event structures with different degrees of integration and that the degree of integration a memory representation achieves may be influenced by moderators such as animacy or agency. Thereby, I provide researchers with a new tool to investigate binding processes in episodic memory and extend a relatively limited body of empirical evidence regarding these binding processes. The somewhat nuanced findings highlight the complexity of the system underlying binding processes in episodic memory and call for the development and refinement of theories and models to better understand this fundamental property of our memory system.

6 Bibliography

Aggleton, J. P., & Brown, M. W. (1999). Episodic memory, amnesia, and the hippocampal-anterior thalamic axis. *Behavioral and Brain Sciences*, *22*(3), 425–444. https://doi.org/10.1017/S0140525X99002034

Alvarez, P., & Squire, L. R. (1994). Memory consolidation and the medial temporal lobe: A simple network model. *Proceedings of the National Academy of Sciences*, *91*(15), 7041–7045. https://doi.org/10.1073/pnas.91.15.7041

Andermane, N., Joensen, B. H., & Horner, A. J. (2021). Forgetting across a hierarchy of episodic representations. *Current Opinion in Neurobiology*, *67*, 50–57. https://doi.org/10.1016/j.conb.2020.08.004

Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(5), 1063–1087. https://doi.org/10.1037/0278-7393.20.5.1063

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, *43*(4), 561–573. https://doi.org/10.1007/bf02293814

Arnold, N. R., Heck, D. W., Bröder, A., Meiser, T., & Boywitt, C. D. (2019). Testing hypotheses about binding in context memory with a hierarchical multinomial modeling approach: A preregistered study. *Experimental Psychology*, *66*(3), 239–251. https://doi.org/10.1027/1618-3169/a000442

Backus, A. R., Bosch, S. E., Ekman, M., Grabovetsky, A. V., & Doeller, C. F. (2016). Mnemonic convergence in the human hippocampus. *Nature Communications*, *7*(1), 11991. https://doi.org/10.1038/ncomms11991

Bakker, A., Kirwan, C. B., Miller, M., & Stark, C. E. L. (2008). Pattern separation in the human hippocampal CA3 and dentate gyrus. *Science*, *319*(5870), 1640–1642. https://doi.org/10.1126/science.1152882

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores.* Addison-Wesley.

Bisby, J. A., Horner, A. J., Bush, D., & Burgess, N. (2018). Negative emotional content disrupts the coherence of episodic memories. *Journal of Experimental Psychology: General*, *147*(2), 243–256. https://doi.org/10.1037/xge0000356

Bonin, P., Gelin, M., & Bugaiska, A. (2014). Animates are better remembered than inanimates: Further evidence from word and picture stimuli. *Memory & Cogni-*

tion, *42*(3), 370–382. https://doi.org/10.3758/s13421-013-0368-8

Bonin, P., Gelin, M., Laroche, B., Méot, A., & Bugaiska, A. (2015). The “how” of animacy effects in episodic memory. *Experimental Psychology*, *62*(6), 371–384. https://doi.org/10.1027/1618-3169/a000308

Boywitt, C. D., & Meiser, T. (2012a). Bound context features are integrated at encoding. *Quarterly Journal of Experimental Psychology*, *65*(8), 1484–1501. https://doi.org/10.1080/17470218.2012.656668

Boywitt, C. D., & Meiser, T. (2012b). The role of attention for context-context binding of intrinsic and extrinsic features. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(4), 1099–1107. https://doi.org/10.1037/a0026988

Brown, R., & Fish, D. (1983). The psychological causality implicit in language. *Cognition*, *14*(3), 237–273. https://doi.org/10.1016/0010-0277(83)90006-9

Burton, R. L., Lek, I., & Caplan, J. B. (2017). Associative independence revisited: Competition between conflicting associations can be resolved or even reversed in one trial. *Quarterly Journal of Experimental Psychology*, *70*(4), 832–857. https://doi.org/10.1080/17470218.2016.1171886

Burton, R. L., Lek, I., Dixon, R. A., & Caplan, J. B. (2019). Associative interference in older and younger adults. *Psychology and Aging*, *34*(4), 558–571. https://doi.org/10.1037/pag0000361

Cai, D. J., Aharoni, D., Shuman, T., Shobe, J., Biane, J., Song, W., Wei, B., Veshkini, M., La-Vu, M., Lou, J., Flores, S. E., Kim, I., Sano, Y., Zhou, M., Baumgaertel, K., Lavi, A., Kamata, M., Tuszynski, M., Mayford, M., … Silva, A. J. (2016). A shared neural ensemble links distinct contextual memories encoded close in time. *Nature*, *534*(7605), 115–118. https://doi.org/10.1038/nature17955

Chen, W.-H., & Thissen, D. (1997). Local dependence indexes for item pairs using item response theory. *Journal of Educational and Behavioral Statistics*, *22*(3), 265–289. https://doi.org/10.3102/10769986022003265

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.

Cohen, N. J., & Eichenbaum, H. (1993). *Memory, amnesia, and the hippocampal system*. MIT Press.

Damasio, A. R. (1989). The brain binds entities and events by multiregional activation from convergence zones. *Neural Computation*, *1*(1), 123–132. https://doi.org/10.1162/neco.1989.1.1.123

Davachi, L., Mitchell, J. P., & Wagner, A. D. (2003). Multiple routes to memory: Distinct medial temporal lobe processes build item and source memories. *Proceedings of the National Academy of Sciences*, *100*(4), 2157–2162. https://doi.org/10.1073/pnas.0337195100

Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their application.* Cambridge University Press.

de Ayala, R. J. (2009). *The theory and practice of item response theory.* Guilford Press.

Debelak, R., & Koller, I. (2020). Testing the local independence assumption of the Rasch model with Q3-based nonparametric model tests. *Applied Psychological Measurement*, *44*(2), 103–117. https://doi.org/10.1177/0146621619835501

DeYoung, C. M., & Serra, M. J. (2021). Judgments of learning reflect the Animacy advantage for memory, but not beliefs about the effect. *Metacognition and Learning.* https://doi.org/10.1007/s11409-021-09264-w

Diana, R. A., Yonelinas, A. P., & Ranganath, C. (2010). Medial temporal lobe activity during source retrieval reflects information type, not memory strength. *Journal of Cognitive Neuroscience*, *22*(8), 1808–1818. https://doi.org/10.1162/jocn.2009.21335

Eichenbaum, H. (1999). The hippocampus and mechanisms of declarative memory. *Behavioural Brain Research*, *103*(2), 123–133. https://doi.org/10.1016/S0166-4328(99)00044-3

Eichenbaum, H. (2017). On the integration of space, time, and memory. *Neuron*, *95*(5), 1007–1018. https://doi.org/10.1016/j.neuron.2017.06.036

Eichenbaum, H., & Cohen, N. J. (1988). Representation in the hippocampus: What do hippocampal neurons code? *Trends in Neurosciences*, *11*(6), 244–248. https://doi.org/10.1016/0166-2236(88)90100-2

Eichenbaum, H., & Cohen, N. J. (2001). *From conditioning to conscious recollection: Memory systems of the brain.* Oxford University Press.

Eichenbaum, H., Yonelinas, A. P., & Ranganath, C. (2007). The medial temporal lobe and recognition memory. *Annual Review of Neuroscience*, *30*, 123–152. https://doi.org/10.1146/annurev.neuro.30.051606.094328

Frings, C., Hommel, B., Koch, I., Rothermund, K., Dignath, D., Giesen, C., Kiesel, A., Kunde, W., Mayr, S., Moeller, B., Möller, M., Pfister, R., & Philipp, A. (2020). Binding and retrieval in action control (BRAC). *Trends in Cognitive Sciences*, *24*(5), 375–387. https://doi.org/10.1016/j.tics.2020.02.004

Gibbons, R. D., & Hedeker, D. R. (1992). Full-information item bi-factor analysis. *Psychometrika*, *57*(3), 423–436. https://doi.org/10.1007/bf02295430

Guzowski, J. F., Knierim, J. J., & Moser, E. I. (2004). Ensemble dynamics of hippocampal regions CA3 and CA1. *Neuron*, *44*(4), 581–584. https://doi.org/10.1016/j.neuron.2004.11.003

Hayman, C. G., & Tulving, E. (1989). Contingent dissociation between recognition and fragment completion: The method of triangulation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(2), 228–240. https://doi.org/10.1037/0278-7393.15.2.228

Healy, M. J., & Caudell, T. P. (2019). Episodic memory: A hierarchy of spatiotemporal concepts. *Neural Networks*, *120*, 40–57. https://doi.org/10.1016/j.neunet.2019.09.021

Hintzman, D. L. (1972). On testing the independence of associations. *Psychological Review*, *79*(3), 261–264. https://doi.org/10.1037/h0032684

Hintzman, D. L. (1980). Simpson's paradox and the analysis of memory retrieval. *Psychological Review*, *87*(4), 398–410. https://doi.org/10.1037/0033-295x.87.4.398

Hitlin, S., & Elder, G. H. (2007). Time, self, and the curiously abstract concept of agency. *Sociological Theory*, *25*(2), 170–191. https://doi.org/10.1111/j.1467-9558.2007.00303.x

Hommel, B. (1998). Event files: Evidence for automatic integration of stimulus-response episodes. *Visual Cognition*, *5*(1-2), 183–216. https://doi.org/10.1080/713756773

Hommel, B. (2004). Event files: Feature binding in and across perception and action. *Trends in Cognitive Sciences*, *8*(11), 494–500. https://doi.org/10.1016/j.tics.2004.08.007

Hommel, B. (2009). Action control according to TEC (theory of event coding). *Psychological Research*, *73*(4), 512–526. https://doi.org/10.1007/s00426-009-0234-2

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The Theory of Event Coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences*, *24*(5), 849–878. https://doi.org/10.1017/s0140525x01000103

Horner, A. J., Bisby, J. A., Bush, D., Lin, W.-J., & Burgess, N. (2015). Evidence for holistic episodic recollection via hippocampal pattern completion. *Nature Communications*, *6*(1), 7462. https://doi.org/10.1038/ncomms8462

Horner, A. J., & Burgess, N. (2013). The associative structure of memory for multielement events. *Journal of Experimental Psychology: General*, *142*(4), 1370–1383. https://doi.org/10.1037/a0033626

Horner, A. J., & Burgess, N. (2014). Pattern completion in multielement event engrams. *Current Biology*, *24*(9), 988–992. https://doi.org/10.1016/j.cub.2014.03.012

Huff, M. J., Meade, M. L., & Hutchison, K. A. (2011). Age-related differences in guessing on free and forced recall tests. *Memory*, *19*(4), 317–330. https://doi.org/10.1080/09658211.2011.568494

Hunt, R. R., & Einstein, G. O. (1981). Relational and item-specific information in memory. *Journal of Verbal Learning and Verbal Behavior*, *20*(5), 497–514. https://doi.org/10.1016/S0022-5371(81)90138-9

James, E., Ong, G., Henderson, L., & Horner, A. J. (2020). Make or break it: Boundary conditions for integrating multiple elements in episodic memory. *Royal Society Open Science*, *7*(9), 200431. https://doi.org/10.1098/rsos.200431

Joensen, B. H., Gaskell, M. G., & Horner, A. J. (2020). United we fall: All-or-none forgetting of complex episodic events. *Journal of Experimental Psychology: General*, *149*(2), 230–248. https://doi.org/10.1037/xge0000648

Johnson, A. H., & Barrett, J. (2003). The role of control in attributing intentional agency to inanimate objects. *Journal of Cognition and Culture*, *3*(3), 208–217. https://doi.org/10.1163/156853703322336634

Kahana, M. J. (2002). Associative symmetry and memory theory. *Memory & Cognition*, *30*(6), 823–840. https://doi.org/10.3758/BF03195769

Kahana, M. J., Rizzuto, D. S., & Schneider, A. R. (2005). Theoretical correlations and measured correlations: Relating recognition and recall in four distributed memory models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(5), 933–953. https://doi.org/10.1037/0278-7393.31.5.933

Kako, E. (2006). Thematic role properties of subjects and objects. *Cognition*, *101*(1), 1–42. https://doi.org/10.1016/j.cognition.2005.08.002

Kasof, J., & Lee, J. Y. (1993). Implicit causality as implicit salience. *Journal of Personality and Social Psychology*, *65*(5), 877–891. https://doi.org/10.1037/0022-3514.65.5.877

Kassin, S., & Lowe, C. (1979). On the use of single sentence descriptions of behavior in attribution research. *Social Behavior and Personality*, *7*(1), 1–8. https://doi.org/10.2224/sbp.1979.7.1.1

Kazanas, S. A., Altarriba, J., & O'Brien, E. G. (2020). Paired-associate learning, animacy, and imageability effects in the survival advantage. *Memory & Cognition*, *48*(2), 244–255. https://doi.org/10.3758/s13421-019-01007-2

Kumaran, D. (2013). Schema-driven facilitation of new hierarchy learning in the transitive inference paradigm. *Learning & Memory*, *20*(7), 388–394. https://doi.org/10.1101/lm.030296.113

Kumaran, D., & Ludwig, H. (2013). Transitivity performance, relational hierarchy knowledge and awareness: Results of an instructional framing manipulation. *Hippocampus*, *23*(12), 1259–1268. https://doi.org/10.1002/hipo.22163

Laurino, J., & Kaczer, L. (2019). Animacy as a memory enhancer during novel word learning: Evidence from orthographic and semantic memory tasks. *Memory*, *27*(6), 820–828. https://doi.org/10.1080/09658211.2019.1572195

Lazarsfeld, P. F., & Henry, N. W. (1968). *Latent structure analysis.* Houghton Mifflin.

Leding, J. K. (2019). Adaptive memory: Animacy, threat, and attention in free recall. *Memory & Cognition*, *47*(3), 383–394. https://doi.org/10.3758/s13421-018-0873-x

Leding, J. K. (2020). Animacy and threat in recognition memory. *Memory & Cognition*, *48*(5), 788–799. https://doi.org/10.3758/s13421-020-01017-5

Li, P., Jia, X., Li, X., & Li, W. (2016). The effect of animacy on metamemory. *Memory & Cognition*, *44*(5), 696–705. https://doi.org/10.3758/s13421-016-0598-7

Lord, F. M. (1980). *Applications of item response theory to practical testing problems.* Erlbaum.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores.* Addison-Wesley.

Lowder, M. W., & Gordon, P. C. (2015). Natural forces as agents: Reconceptualizing the animate–inanimate distinction. *Cognition*, *136*, 85–90. https://doi.org/10.1016/j.cognition.2014.11.021

Madan, C. R. (2021). Exploring word memorability: How well do different word properties explain item free-recall probability? *Psychonomic Bulletin & Review*, *28*(2), 583–595. https://doi.org/10.3758/s13423-020-01820-w

Marr, D. (1971). Simple memory: A theory for archicortex. *Philosophical Transactions of the Royal Society of London. B: Biological Sciences*, *262*(841), 23–81. https://doi.org/10.1098/rstb.1971.0078

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, *47*(2), 149–174. https://doi.org/10.1007/bf02296272

McClelland, J. L., O'Reilly, R. C., & McNaughton, B. L. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*(3), 419–457. https://doi.org/10.1037/0033-295x.102.3.419

Meiser, T., & Bröder, A. (2002). Memory for multidimensional source information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(1), 116–137. https://doi.org/10.1037//0278-7393.28.1.116

Moeller, B., Pfister, R., Kunde, W., & Frings, C. (2019). Selective binding of stimulus, response, and effect features. *Psychonomic Bulletin & Review*, *26*(5), 1627–1632. https://doi.org/10.3758/s13423-019-01646-1

Moll, M., & Miikkulainen, R. (1997). Convergence-zone episodic memory: Analysis and simulations. *Neural Networks*, *10*(6), 1017–1036. https://doi.org/10.1016/S0893-6080(97)00016-6

Morton, N. W., Schlichting, M. L., & Preston, A. R. (2020). Representations of common event structure in medial temporal lobe and frontoparietal cortex support efficient inference. *Proceedings of the National Academy of Sciences*, *117*(47), 29338–29345. https://doi.org/10.1073/pnas.1912338117

Nairne, J. S., Pandeirada, J. N. S., & Thompson, S. R. (2008). Adaptive memory: The comparative value of survival processing. *Psychological Science*, *19*(2), 176–180. https://doi.org/10.1111/j.1467-9280.2008.02064.x

Nairne, J. S., Thompson, S. R., & Pandeirada, J. N. S. (2007). Adaptive memory: Survival processing enhances retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(2), 263–273. https://doi.org/10.1037/0278-7393.33.2.263

Nairne, J. S., VanArsdall, J. E., & Cogdill, M. (2017). Remembering the living: Episodic memory is tuned to animacy. *Current Directions in Psychological Science*, *26*(1), 22–27. https://doi.org/10.1177/0963721416667711

Nairne, J. S., VanArsdall, J. E., Pandeirada, J. N. S., Cogdill, M., & LeBreton, J. M. (2013). Adaptive memory: The mnemonic value of animacy. *Psychological Science*, *24*(10), 2099–2105. https://doi.org/10.1177/0956797613480803

Nakazawa, K., Quirk, M. C., Chitwood, R. A., Watanabe, M., Yeckel, M. F., Wilson, M. A., & Tonegawa, S. (2002). Requirement for hippocampal CA3 NMDA receptors in associative memory recall. *Science*, *297*(5579), 211–218. https://doi.org/10.1126/science.1071795

Naveh-Benjamin, M. (2000). Adult age differences in memory performance: Tests of an associative deficit hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(5), 1170–1187. https://doi.org/10.1037/0278-7393.26.5.1170

Neunuebel, J. P., & Knierim, J. J. (2014). CA3 retrieves coherent representations from degraded input: Direct evidence for CA3 pattern completion and dentate gyrus pattern separation. *Neuron*, *81*(2), 416–427. https://doi.org/10.1016/j.neuron.2013.11.017

Ngo, C. T., Horner, A. J., Newcombe, N. S., & Olson, I. R. (2019). Development of holistic episodic recollection. *Psychological Science*, *30*(12), 1696–1706. https://doi.org/10.1177/0956797619879441

Ngo, C. T., & Newcombe, N. S. (2021). Relational binding and holistic retrieval in ageing. *Memory*, *29*(9), 1197–1205. https://doi.org/10.1080/09658211.2021.1974047

O'Mara, S. M. (2005). The subiculum: What it does, what it might do, and what neuroanatomy has yet to tell us. *Journal of Anatomy*, *207*(3), 271–282. https://doi.org/10.1111/j.1469-7580.2005.00446.x

O'Mara, S. M., Commins, S., Anderson, M., & Gigg, J. (2001). The subiculum: A review of form, physiology and function. *Progress in Neurobiology*, *64*(2), 129–155. https://doi.org/10.1016/S0301-0082(00)00054-X

Old, S. R., & Naveh-Benjamin, M. (2008). Differential effects of age on item and associative measures of memory: A meta-analysis. *Psychology and Aging*, *23*(1), 104–118. https://doi.org/10.1037/0882-7974.23.1.104

Ponocny, I. (2001). Nonparametric goodness-of-fit tests for the rasch model. *Psychometrika*, *66*(3), 437–459. https://doi.org/10.1007/BF02294444

Popp, E. Y., & Serra, M. J. (2016). Adaptive memory: Animacy enhances free recall but impairs cued recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(2), 186–201. https://doi.org/10.1037/xlm0000174

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests.* Nielsen & Lydiche.

Reese, E., Haden, C. A., Baker-Ward, L., Bauer, P., Fivush, R., & Ornstein, P. A. (2011). Coherence of personal narratives across the lifespan: A multidimensional model and coding method. *Journal of Cognition and Development*, *12*(4), 424–462. https://doi.org/10.1080/15248372.2011.587854

Robin, J. (2018). Spatial scaffold effects in event memory and imagination. *WIREs Cognitive Science*, *9*(4), e1462. https://doi.org/10.1002/wcs.1462

Rolls, E. T. (2016). Pattern separation, completion, and categorisation in the hippocampus and neocortex. *Neurobiology of Learning and Memory*, *129*, 4–28. https://doi.org/10.1016/j.nlm.2015.07.008

Rubin, D. C., & Umanath, S. (2015). Event memory: A theory of memory for laboratory, autobiographical, and fictional events. *Psychological Review*, *122*(1), 1–23. https://doi.org/10.1037/a0037907

Saunders, R. C., & Rosene, D. L. (1988). A comparison of the efferents of the amygdala and the hippocampal formation in the rhesus monkey: I. Convergence in the entorhinal, prorhinal, and perirhinal cortices. *Journal of Comparative Neurology*, *271*(2), 153–184. https://doi.org/10.1002/cne.902710202

Schapiro, A. C., Turk-Browne, N. B., Botvinick, M. M., & Norman, K. A. (2017). Complementary learning systems within the hippocampus: A neural network modelling approach to reconciling episodic memory with statistical learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1711), 20160049. https://doi.org/10.1098/rstb.2016.0049

Schlichting, M. L., Guarino, K. F., Schapiro, A. C., Turk-Browne, N. B., & Preston, A. R. (2017). Hippocampal structure predicts statistical learning and associative inference abilities during development. *Journal of Cognitive Neuroscience*, *29*(1), 37–51. https://doi.org/10.1162/jocn_a_01028

Schlichting, M. L., Zeithamova, D., & Preston, A. R. (2014). CA_1 subfield contributions to memory integration and inference. *Hippocampus*, *24*(10), 1248–1260. https://doi.org/10.1002/hipo.22310

Schreiner, M. R., Bröder, A., & Meiser, T. (2022). *Agency effects on the binding of event elements in episodic memory.* Manuscript submitted for publication.

Schreiner, M. R., & Meiser, T. (2022). Measuring binding effects in event-based episodic representations. *Behavior Research Methods.* Advance online publication. https://doi.org/10.3758/s13428-021-01769-1

Schreiner, M. R., Meiser, T., & Bröder, A. (2022). The binding structure of event elements in episodic memory and the role of animacy. *Quarterly Journal of Experimental Psychology.* Advance online publication. https://doi.org/10.1177/17470218221096148

Serra, M. J., & DeYoung, C. M. (2022). Within-pair factors might explain the inconsistent effects of animacy on paired-associates recall. *Psychonomic Bulletin & Review.* https://doi.org/10.3758/s13423-022-02184-z

Shohamy, D., & Wagner, A. D. (2008). Integrating memories in the human brain:

Hippocampal-midbrain encoding of overlapping events. *Neuron*, *60*(2), 378–389. https://doi.org/10.1016/j.neuron.2008.09.023

Simpson, E. H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society*, *13*(2), 238–241. https://doi.org/10.1111/j.2517-6161.1951.tb00088.x

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, *117*(1), 34–50. https://doi.org/10.1037/0096-3445.117.1.34

Squire, L., & Zola-Morgan, S. (1991). The medial temporal lobe memory system. *Science*, *253*(5026), 1380–1386. https://doi.org/10.1126/science.1896849

Starns, J. J., & Hicks, J. L. (2005). Source dimensions are retrieved independently in multidimensional monitoring tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(6), 1213–1220. https://doi.org/10.1037/0278-7393.31.6.1213

Starns, J. J., & Hicks, J. L. (2008). Context attributes in memory are bound to item information, but not to one another. *Psychonomic Bulletin & Review*, *15*(2), 309–314. https://doi.org/10.3758/PBR.15.2.309

Sugar, J., & Moser, M.-B. (2019). Episodic memory: Neuronal codes for what, where, and when. *Hippocampus*, *29*(12), 1190–1205. https://doi.org/10.1002/hipo.23132

Suitner, C., & Maass, A. (2016). Spatial agency bias: Representing people in space. In J. M. Olson & M. P. Zanna (Eds.), *Advances in Experimental Social Psychology* (Vol. 53, pp. 245–301). Academic Press. https://doi.org/10.1016/bs.aesp.2015.09.004

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory* (pp. 381–403). Academic Press.

Tulving, E. (1983). *Elements of episodic memory*. Oxford University Press.

Tulving, E. (1993). What is episodic memory? *Current Directions in Psychological Science*, *2*(3), 67–70. https://doi.org/10.1111/1467-8721.ep10770899

van Strien, N. M., Cappaert, N. L. M., & Witter, M. P. (2009). The anatomy of memory: An interactive overview of the parahippocampal–hippocampal network. *Nature Reviews Neuroscience*, *10*(4), 272–282. https://doi.org/10.1038/nrn2614

VanArsdall, J. E., Nairne, J. S., Pandeirada, J. N. S., & Blunt, J. R. (2013). Adaptive memory: Animacy processing produces mnemonic advantages. *Experimental Psychology*, *60*(3), 172–178. https://doi.org/10.1027/1618-3169/a000186

VanArsdall, J. E., Nairne, J. S., Pandeirada, J. N. S., & Cogdill, M. (2015). Adaptive memory: Animacy effects persist in paired-associate learning. *Memory*, *23*(5), 657–663. https://doi.org/10.1080/09658211.2014.916304

Verhelst, N. D. (2008). An efficient MCMC algorithm to sample binary matrices with fixed marginals. *Psychometrika*, *73*(4), 705–728. https://doi.org/10.1007/s11336-008-9062-3

Wainer, H., & Wang, X. (2000). Using a new statistical model for testlets to score TOEFL. *Journal of Educational Measurement*, *37*(3), 203–220. https://doi.org/10.1111/j.1745-3984.2000.tb01083.x

Yassa, M. A., & Stark, C. E. L. (2011). Pattern separation in the hippocampus. *Trends in Neurosciences*, *34*(10), 515–525. https://doi.org/10.1016/j.tins.2011.06.006

Yen, W. M. (1984). Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Applied Psychological Measurement*, *8*(2), 125–145. https://doi.org/10.1177/014662168400800201

Yen, W. M. (1993). Scaling performance assessments: Strategies for managing local item dependence. *Journal of Educational Measurement*, *30*(3), 187–213. https://doi.org/10.1111/j.1745-3984.1993.tb00423.x

Yule, G. U. (1912). On the methods of measuring association between two attributes. *Journal of the Royal Statistical Society*, *75*(6), 579–652. https://doi.org/10.2307/2340126

Zeithamova, D., Schlichting, M. L., & Preston, A. R. (2012). The hippocampus and inferential reasoning: Building memories to navigate future decisions. *Frontiers in Human Neuroscience*, *6*. https://doi.org/10.3389/fnhum.2012.00070

Zotow, E., Bisby, J. A., & Burgess, N. (2020). Behavioral evidence for pattern separation in human episodic memory. *Learning & Memory*, *27*(8), 301–309. https://doi.org/10.1101/lm.051821.120

A Acknowledgements

"I can't carry it for you, but I can carry you!"

from *The Return of the King*
by J. R. R. Tolkien

I would like to thank all the people who supported me during my dissertation. You have my sincere gratitude.

First, I would like to thank Thorsten and Arndt for their invaluable advice, guidance, and encouragement, while giving me a lot of flexibility in pursuing this dissertation. I thank Mandy for providing helpful advice along the way. I further thank Ali, Neal, Christine, Nicole, Alex, and Joel for hosting me in Austin and providing me with new perspectives on my projects and helpful feedback.

Many thanks to Maren and Julian for their valuable advice and feedback. I further like to thank Tobias, Barbara, Annika, Gloria, my lab mates Nils, Franziska, Susanne, Viola, Fabiola, Julia, and all of SMiP for the insightful discussions, enjoyable coffee and lunch breaks, informal meetings, retreats, conferences, and stimulating workshops we spent together.

I am particularly grateful for the support of my girlfriend Anne, who always supported and encouraged me in both difficult and good times, while also being a great discussion partner and providing me with helpful advice and feedback.

Finally, I would like to thank all my family and friends, particularly my parents Carola and Franz, for their encouragement and support throughout my dissertation. Thank you.

B Statement of Originality

1. I hereby declare that the presented doctoral dissertation with the title *Binding Processes in Episodic Memory: Measurement, Structure, and Moderators* is my own work.
2. I did not seek unauthorized assistance of a third party and I have employed no other sources or means except the ones listed. I clearly marked any quotations derived from the works of others.
3. I did not yet present this doctoral dissertation or parts of it at any other higher education institution in Germany or abroad.
4. I hereby confirm the accuracy of the declaration above.
5. I am aware of the significance of this declaration and the legal consequences in case of untrue or incomplete statements.

I affirm in lieu of oath that the statements above are to the best of my knowledge true and complete.

Signature:

Date:

C Co-Authors' Statements

Co-Author: Thorsten Meiser

With this statement, I confirm that the following articles included in the present thesis were primarily conceived and written by Marcel R. Schreiner.

Schreiner, M. R., & Meiser, T. (2022). Measuring binding effects in event-based episodic representations. *Behavior Research Methods*. Advance online publication. https://doi.org/10.3758/s13428-021-01769-1

Schreiner, M. R., Meiser, T., & Bröder, A. (2022). The binding structure of event elements in episodic memory and the role of animacy. *Quarterly Journal of Experimental Psychology*. Advance online publication. https://doi.org/10.1177/17470218221096148

Schreiner, M. R., Bröder, A., & Meiser, T. (2022). *Agency effects on the binding of event elements in episodic memory.* Manuscript submitted for publication.

Marcel R. Schreiner designed and conducted all experiments reported in Schreiner, Meiser, and Bröder (2022) and Schreiner, Bröder, and Meiser (2022) as well as the simulation reported in Schreiner and Meiser (2022). He analyzed all data in all three manuscripts and wrote most of the articles, including first drafts and revisions of the manuscripts. I contributed to the development and refinement of the research questions, study design, and statistical analyses. Furthermore, I revised the manuscripts.

Prof. Dr. Thorsten Meiser
Mannheim, October 2022

Co-Author: Arndt Bröder

With this statement, I confirm that the following articles included in the present thesis were primarily conceived and written by Marcel R. Schreiner.

Schreiner, M. R., Meiser, T., & Bröder, A. (2022). The binding structure of event elements in episodic memory and the role of animacy. *Quarterly Journal of Experimental Psychology.* Advance online publication. https://doi.org/10.1177/17470218221096148

Schreiner, M. R., Bröder, A., & Meiser, T. (2022). *Agency effects on the binding of event elements in episodic memory.* Manuscript submitted for publication.

Marcel R. Schreiner designed and conducted all experiments reported in Schreiner, Meiser, and Bröder (2022) and Schreiner, Bröder, and Meiser (2022). He analyzed all data in both manuscripts and wrote most of the articles, including first drafts and revisions of the manuscripts. I contributed to the development and refinement of the research questions and study design. Furthermore, I revised the manuscripts.

Prof. Dr. Arndt Bröder
Mannheim, October 2022

D Copies of Articles

Behavior Research Methods
https://doi.org/10.3758/s13428-021-01769-1

Measuring binding effects in event-based episodic representations

Marcel R. Schreiner[1] · **Thorsten Meiser**[1]

Accepted: 8 December 2021

Abstract

Remembering an experienced event in a coherent manner requires the binding of the event's constituent elements. Such binding effects manifest as a stochastic dependency of the retrieval of event elements. Several approaches for modeling these dependencies have been proposed. We compare the contingency-based approach by Horner & Burgess (*Journal of Experimental Psychology: General, 142*(4), 1370–1383, 2013), related approaches using Yule's Q (Yule, *Journal of the Royal Statistical Society, 75*(6), 579–652, 1912) or an adjusted Yule's Q (c.f. Horner & Burgess, *Current Biology, 24*(9), 988–992, 2014), an approach based on item response theory (IRT, Schreiner et al., in press), and a nonparametric variant of the IRT-based approach. We present evidence from a simulation study comparing the five approaches regarding their empirical detection rates and susceptibility to different levels of memory performance, and from an empirical application. We found the IRT-based approach and its nonparametric variant to yield the highest power for detecting dependencies or differences in dependency between conditions. However, the nonparametric variant yielded increasing Type I error rates with increasing dependency in the data when testing for differences in dependency. We found the approaches based on Yule's Q to yield biased estimates and to be strongly affected by memory performance. The other measures were unbiased given no dependency or differences in dependency but were also affected by memory performance if there was dependency in the data or if there were differences in dependency, but to a smaller extent. The results suggest that the IRT-based approach is best suited for measuring binding effects. Further considerations when deciding for a modeling approach are discussed.

Keywords Statistical modeling · Episodic memory · Binding · Item response theory

Storing information about experienced events in episodic memory requires the events' constituent elements to be bound together. Such binding processes allow for a coherent retrieval of the experienced event. An event's constituent elements may take very different forms such as persons, objects, locations, actions, and sensations. For example, imagine having bought bread at a bakery. Later remembering this particular event requires different elements such as the bakery (location), the bought bread (object), and the vendor (person) to be bound together in memory. If event elements are bound together, there should be an increased likelihood of retrieving subsequent events elements when a preceding element was successfully retrieved, thus leading to a stochastic dependency of the retrieval of event elements (e.g., Arnold et al., 2019; Boywitt & Meiser, 2012a, b; Horner et al., 2015; Horner & Burgess, 2013, 2014; Meiser and Bröder, 2002; Ngo et al., 2019; Starns & Hicks, 2005, 2008).

Much of the past research on binding in episodic memory (e.g., Balaban et al., 2019; Boywitt & Meiser, 2012a, b; Hicks and Starns, 2016; Meiser and Bröder, 2002; Starns & Hicks, 2005, 2008; Utochkin and Brady, 2020; Vogt and Bröder, 2007) investigated rather simple, item-based representations. Item-based representations consist of a single element with specific features, such as an object with a certain shape or color. Thus, item-based representations are static (see also Hunt & Einstein, 1981). More recently, research started to incorporate more complex, event-based representations that may include several elements (e.g., Andermane et al., 2021, Horner et al., 2015, 2013, 2014; James et al., 2020, Joensen et al., 2020). These elements may interact and thus, event-based representations are, at least potentially, dynamic (see also Rubin & Umanath, 2015). In this context, the presentation of different elements belonging to the same event may induce relational encoding with features common to the same event (Hunt & Einstein, 1981). Event-based representations can

✉ Marcel R. Schreiner
m.schreiner@uni-mannheim.de

[1] Department of Psychology, School of Social Sciences, University of Mannheim, L13, 15, 68161 Mannheim, Germany

be considered to contain several item-based representations, with storage occuring in a hierarchical manner (i.e., item-based representations being nested in event-based representations, see Andermane et al., 2021) or event- and item-based representations can be distinguished based on different degrees of discrimination, with item-based representations containing more specific information than event-based representations (Hunt & Einstein, 1981). Additionally, event-based representations include a spatiotemporal context, which is not the case for item-based representations (e.g., Andermane et al., 2021). Contrary to item-based representations, event-based representations allow for the construction of scenes (Robin, 2018; Rubin & Umanath, 2015). This scene construction does not necessitate the exact remembering of the specific features of an event's constituent elements (Rubin & Umanath, 2015). Most research on event-based representations has not considered specific features of the events' constituent elements, which however have been a main focus of research on item-based representations (e.g., Balaban et al., 2019; Horner and Burgess, 2013; Joensen et al., 2020; Utochkin and Brady., 2020).

Because event-based representations are more complex than item-based representations, approaches for modeling stochastic dependencies of the retrieval of event elements developed for item-based representations can not be readily applied to event-based representations. Instead, different approaches have been proposed for event-based representations. The different approaches are first introduced before reporting a simulation study comparing the approaches regarding their power for detecting stochastic dependency of the retrieval of event elements and differences in dependency, Type I error rates, and susceptibility to variations in memory performance. The approaches are then applied to an empirical data example to evaluate the congruence of empirical inferences drawn by using the different approaches.

Approach by Horner and Burgess

Horner and Burgess (2013) proposed a contingency-based approach that can be applied to data obtained from cued recognition or cued recall tasks. The approach considers items (i.e., test trials in a memory test) with a common cue or target as a dependency pair. For example, if events consist of the elements A, B, and C, the cue-target-pairs A–B and A–C may be considered a dependency pair. For each person *i*, event *t*, and dependency pair *jj'* a contingency table **X** showing the successful retrieval of the target of a dependency pair can be constructed, with 1 denoting successful retrieval and 0 a failure to retrieve the target:

$$\mathbf{X}_{\mathrm{it}}^{\mathrm{jj'}} = \begin{bmatrix} j=1, j'=1 & j=1, j'=0 \\ j=0, j'=1 & j=0, j'=0 \end{bmatrix} \tag{1}$$

Summing over events, a contingency table for a given person and dependency pair can be obtained:

$$\mathbf{X}_{\mathrm{i}}^{\mathrm{jj'}} = \begin{bmatrix} n_{11} & n_{10} \\ n_{01} & n_{00} \end{bmatrix} \tag{2}$$

n_{11} is the frequency of both items of a dependency pair being correctly retrieved across events, n_{10} is the frequency of item *j* being correctly retrieved while item *j'* being incorrectly retrieved, n_{01} is the frequency of item *j* being incorrectly retrieved while item *j'* being correctly retrieved, and n_{00} is the frequency of both items being incorrectly retrieved. From these contingency tables (one per dependency pair), Horner and Burgess (2013) calculate a data-based measure of the dependency of the retrieval of event elements. The measure is first calculated for each dependency pair by summing the leading diagonal cells of each contingency table per person and dividing the results by the overall number of events *T*. Then the results are averaged across the set of dependency pairs *J*:

$$D_{\mathrm{HB,\,i}}^{\mathrm{data}} = \frac{1}{|J|} \sum_{jj' \in J} \frac{n_{11} + n_{00}}{T} \tag{3}$$

The measure reflects the mean proportion of items in an event that were both successfully or unsuccessfully retrieved. Because this measure necessarily increases if many (or few) event elements are successfully retrieved due to strong (or poor) overall memory performance, Horner and Burgess (2013) contrast it with dependency estimates from an "independent model," which predicts a value of the measure under the assumption of independence based on the person's mean performance for items of a dependency pair across events:

$$D_{\mathrm{HB,\,i}}^{\mathrm{ind}} = \frac{1}{|J|} \sum_{jj' \in J} \left(\frac{n_{11} + n_{10}}{T} \frac{n_{11} + n_{01}}{T} + \left(1 - \frac{n_{11} + n_{10}}{T}\right)\left(1 - \frac{n_{11} + n_{01}}{T}\right) \right) \tag{4}$$

The actual dependency measure $D_{\mathrm{HB,\,i}}$ can then be obtained by subtracting $D_{\mathrm{HB,\,i}}^{\mathrm{ind}}$ from $D_{\mathrm{HB,\,i}}^{\mathrm{data}}$. The measure can take values between -1 and 1. A value of 0 indicates independence, positive values indicate dependency, and negative values indicate negative dependency such that the likelihood of retrieving an event element is smaller when a preceding event element was successfully retrieved.

Yule's Q

Similarly to the approach by Horner and Burgess (2013), one can calculate a measure of dependency from the contingency table in Eq. 2 using Yule's Q (Yule, 1912; cf. Horner and Burgess, 2014; see also Hayman and Tulving, 1989; Kahana, 2002; Kahana et al., 2005), a commonly used measure of association in memory research. Yule's Q is an odds ratio standardized to the value range of [-1, 1] with the same interpretation as the dependency measure by Horner and Burgess (2013). It is a special case of the gamma coefficient (Goodman & Kruskal, 1954) for 2 × 2 matrices and can be calculated as:

$$Q_{\mathrm{i}}^{\mathrm{jj'}} = \frac{n_{11}n_{00} - n_{10}n_{01}}{n_{11}n_{00} + n_{10}n_{01}} \tag{5}$$

As in the approach by Horner and Burgess (2013), one can then average across dependency pairs:

$$Q_{\mathrm{i}} = \frac{1}{|J|} \sum_{jj' \in J} Q_{\mathrm{i}}^{\mathrm{jj'}} \tag{6}$$

Adjusted Yule's Q

A known problem of Yule's Q is that zero frequencies cause it to become -1, 1, or undefined. One can circumvent this problem by adding a constant such as 0.5 to each cell of the contingency table in Eq. 2 (cf. Burton et al., 2019; Horner and Burgess, 2014). One can then calculate the adjusted Yule's Q (Q_{a}) as in Eqs. 5 and 6. However, as opposed to the approach by Horner and Burgess (2013), the approaches involving Yule's Q do not attempt to correct for memory performance.

All the approaches mentioned so far are contingency-based, collapsing smaller contingency tables into a 2 × 2 contingency table per participant and dependency pair. Thus, the approaches may be prone to Simpson's paradox (Hintzman, 1972, 1980; Simpson, 1951), meaning that if 2 × 2 contingency tables are collapsed into a summary one, the relationship of the two outcomes in the summary table may differ from the one shown in any of the original tables. This may occur due to confounding with participant differences, item differences, or participant-item interactions (Hintzman, 1972, 1980; see also Burton et al., 2017). Since the approaches compute participant-specific estimates, the problem of confounding with participant differences is avoided. However, potential confounding with item differences and, most notably, participant-item interactions remains an issue. Consequently, the approaches for estimating dependency using contingency analyses may be subject to problems of confounding.

An IRT-based approach

Recently, Schreiner et al. (in press) proposed a measure of the retrieval of event elements based on item response theory (IRT, Lord, 1980; Lord and Novick, 1968). Contrary to the approaches outlined before, this measure is not contingency-based but operates on the level of individual item responses (i.e., test trial outcomes in a memory test). Thus, Simpson's paradox does not apply. In addition, IRT jointly models participant differences, item differences, and participant-item interactions, thus avoiding confounding with these covariates. By using the three-parameter logistic model (Birnbaum, 1968), one can model the probability of person *i* to give a correct response *u* to item *j*, given a latent trait θ, which represents memory performance in the current application of the model, an item difficulty β, an item-specific discrimination parameter α, and an item-specific guessing parameter γ:

$$P(u_{\mathrm{ij}} = 1) = \gamma_{\mathrm{j}} + (1 - \gamma_{\mathrm{j}}) \frac{e^{\alpha_{\mathrm{j}}(\theta_{\mathrm{i}} - \beta_{\mathrm{j}})}}{1 + e^{\alpha_{\mathrm{j}}(\theta_{\mathrm{i}} - \beta_{\mathrm{j}})}} \tag{7}$$

In experimental settings, events are often randomly generated. Thus, it is often appropriate to fix the discrimination and guessing parameters. For example, when using cued recognition tests, it may be appropriate to fix the guessing parameter to the stochastic guessing probability derived from the number of response options (e.g., 0.2 for five response options). Discrimination parameters may be fixed to 1, as is the case in the Rasch model (Rasch, 1960), assuming all items having the same correlation or factor loading with the latent trait. When fixing the discrimination parameters to 1 and the guessing parameters to a constant *g*, the model is reduced to:

$$P(u_{\mathrm{ij}} = 1) = g + (1 - g) \frac{e^{\theta_{\mathrm{i}} - \beta_{\mathrm{j}}}}{1 + e^{\theta_{\mathrm{i}} - \beta_{\mathrm{j}}}} \tag{8}$$

This model assumes local independence (LI) of item responses, which means that all inter-item relationships are accounted for by the latent trait (de Ayala, 2009; Lazarsfeld and Henry, 1968). If the LI assumption holds, item residual correlations are zero. However, when binding of event elements occurs there are additional event-specific effects that violate the LI assumption. Consequently, item-residual correlations within events deviate from zero. Item-residual correlations can be estimated using the Q_3 statistic (Yen, 1984), which is calculated for item pairs *jj'* in four steps: First, person and item parameters are estimated from the model in Eqs. 7 or 8. Second, the probability of correctly retrieving items *j* and *j'* is predicted from the model parameters. Third, the residuals for both items are calculated by subtracting the model-implied probability of a correct response from the observed response for each person. Finally, Q_3 is calculated as the correlation of the residuals of both items. The Q_3 statistic has an expected value of $\frac{-1}{I-1}$ given LI, with *I* being the total number of items (Yen, 1993). Thus, Q_3 is negatively biased and in an additional step a bias correction should be applied by subtracting the expected value from all Q_3. Schreiner et al. (in press) then constructed a measure of the dependency of the retrieval of event elements (D_{Q_3}) as the difference in mean within-event and mean between-event Q_3:

$$D_{\mathrm{Q}_3} = \frac{1}{K}\sum_{k>k'} Q_3^{\mathrm{kk}'} - \frac{1}{L}\sum_{l>l'} Q_3^{\mathrm{ll}'} \tag{9}$$

where *kk'* are within-event item pairs, *ll'* are between-event item pairs, K is the total number of within-event item pairs and L is the total number of between-event item pairs. Given binding of event elements, within-event residual correlations deviate from zero and between-event residual correlations are close to zero. Consequently, D_{Q_3} deviates from zero. Like D_{HB} and Yule's Q, D_{Q_3} can take values between -1 and 1 and its interpretation is equivalent to the former measures.

Because the sampling distribution of Q_3, and consequently the one of D_{Q_3}, is unknown (Chen & Thissen, 1997) and D_{Q_3} is an overall, not person-specific, measure of dependency, testing the dependency by means of *t*-tests or linear mixed models, which can be applied to the contingency-based approaches, is not possible. Instead, parametric bootstrapping can be applied, which is a simulation-based approach to generate data from estimated parameters to simulate a distribution of a statistic under the assumption that the data-generating model is true. There are generally two tests that are of interest: testing whether dependency is different from zero and testing whether dependency differs between experimental conditions or groups. For the first test, artificial response matrices can be repeatedly sampled from the model in Eq. 8, with item parameters and latent trait variance estimated from the original response matrix. For each simulated sample one can then calculate D_{Q_3} to obtain distributions under the null hypothesis of independence. From these distributions one can then calculate p values for the observed D_{Q_3}. For the second test, the parametric bootstrap requires estimates of the event-specific effects, which can be obtained by fitting a bifactor model (see Gibbons and Hedeker, 1992; Wainer and Wang, 2000). This model extends the model in Eq. 7 by including additional event-specific latent traits λ:

$$P(u_{\mathrm{ij}} = 1) = \gamma_{\mathrm{j}} + (1-\gamma_{\mathrm{j}})\frac{e^{\alpha_{\mathrm{j}}(\theta_{\mathrm{i}}-\beta_{\mathrm{j}})-\alpha_{\mathrm{t(j)}}\lambda_{\mathrm{it(j)}}}}{1+e^{\alpha_{\mathrm{j}}(\theta_{\mathrm{i}}-\beta_{\mathrm{j}})-\alpha_{\mathrm{t(j)}}\lambda_{\mathrm{it(j)}}}} \tag{10}$$

with λ being the event-specific latent trait of person i for event *t(j)* to which item j belongs. When applying the same restrictions as in Eq. 8 the model reduces to:

$$P(u_{\mathrm{ij}} = 1) = g + (1-g)\frac{e^{\theta_{\mathrm{i}}-\beta_{\mathrm{j}}-\lambda_{\mathrm{it(j)}}}}{1+e^{\theta_{\mathrm{i}}-\beta_{\mathrm{j}}-\lambda_{\mathrm{it(j)}}}} \tag{11}$$

All latent traits in this model are mutually independent. The event-specific latent traits exert their influence via their variance. Higher variances indicate stronger event-specific effects. In experimental settings this model requires an additional latent trait for each event and thus quickly becomes very high-dimensional. It is thus advisable to put equality constraints on the event-specific trait variances within experimental conditions. Using the estimates of latent trait variances and item parameters one can then repeatedly sample artificial response matrices from the model in Eq. 11, while setting the latent trait variances equal to the ones of a given experimental condition (a reference condition). For example, when having two experimental conditions, one may set the latent trait variance of the second condition equal to the one of the first condition, making the model assume no difference in dependency between conditions. One can then calculate D_{Q_3} for each experimental condition and differences in D_{Q_3} between conditions to obtain distributions under the null hypothesis of equal dependency between conditions relative to the reference condition. From these distributions one can then calculate p values for the observed differences in D_{Q_3}.

Nonparametric variant of the IRT-based approach

While the previously presented IRT-based approach (Schreiner et al., in press) is parametric and requires the estimation of item and person parameters, Debelak and Koller (2020) recently proposed a nonparametric estimation procedure for Q_3, building on the nonparametric testing framework by Ponocny (2001). Using a Markov-Chain Monte-Carlo algorithm by Verhelst (2008), a bootstrap sample of artifical response matrices with the same marginal sums as the original response matrix is generated. In the Rasch model (Rasch, 1960), and also the restricted model in Eq. 8, the marginal person sums are sufficient statistics for the general latent trait. It is then possible to estimate $P(u_{\mathrm{ij}} = 1)$ by averaging u_{ij} over all bootstrap samples. The nonparametric variant of Q_3 is then computed like its parametric counterpart, using the estimated $P(u_{\mathrm{ij}} = 1)$ as the model-implied probability of a correct response. Based on the obtained nonparametric variants of Q_3 one can then calculate a dependency measure ($D^{\mathrm{np}}_{\mathrm{Q}_3}$) as in Eq. 9. Similarly as in the parametric approach it is then also possible to calculate $D^{\mathrm{np}}_{\mathrm{Q}_3}$ for each bootstrap sample and to calculate p values for $D^{\mathrm{np}}_{\mathrm{Q}_3}$ and differences in $D^{\mathrm{np}}_{\mathrm{Q}_3}$.

Desirable properties for measures of binding effects in episodic memory are: high power in detecting stochastic dependency of the retrieval of event elements and differences in dependency, good maintenance of Type I error rates, and non-sensitivity to variations in memory performance. Type I error rates and power are central concepts for statistical hypothesis testing (see e.g., Cohen, 1988) in order to guarantee strict statistical tests and replicable findings. In addition, binding effects should be dissociated from memory performance, which requires measures of binding effects that are unaffected by memory performance, because otherwise it is unclear whether increased dependency of the retrieval of event elements can be attributed to actual binding effects or is due to higher levels of memory performance in the sample, which also increases the likelihood that several elements from the same event are correctly retrieved. In a simulation study we compared the five presented approaches regarding these criteria.

Simulation study

Methods

We conducted a Monte Carlo simulation. Responses were generated from the bifactor model in Eq. 11 with a global guessing parameter of $g = 0.2$, $t = 30$ events, and 6 items (i.e., test trials in a hypothetical memory test) per event, resulting in a total of $I = 180$ items. In an application, this scenario could be equivalent to testing each association of events consisting of three elements A, B, and C in both directions using a cued recognition task (i.e. testing the cue-target pairs A–B, B–A, A–C, C–A, B–C, and C–B). The different test trials represent the items. The simulation mimicked 2 experimental within-subjects conditions, resulting in 15 events and 90 items per experimental condition.

Item parameters were drawn from a standard normal distribution. Person parameters (i.e., latent memory proficiency [θ] and event-specific latent trait scores [λ_t]) were drawn from a multivariate normal distribution with zero covariances, since the bifactor model assumes the general and event-specific latent traits to be mutually independent (e.g., Wang & Wilson, 2005). The mean of the general latent trait, representing overall memory performance, varied across simulation conditions the and variance was set to 5, based on empirical findings (cf. Schreiner et al., in press). The means of the event-specific latent traits were set to zero and the variances varied across simulation conditions. Variances were constrained to be equal within experimental conditions.

There were four design factors in the simulation: (a) sample size ($N = \{25, 50, 75, 100\}$), (b) dependency (event-specific trait variances, Dep. = {0, 0.5, 1}), (c) differences in dependency (differences in event-specific trait variances, $\text{Dep.}_{\text{diff}} = \{0, 0.5, 1\}$), and (d) overall level of memory performance (mean of the general latent trait θ, $P = \{-2, 0, 2\}$). Different levels of memory performance resulted in proportions of 40%-42% ($P = -2$), 59%-60% ($P = 0$), and 75%-80% ($P = 2$) correct responses. The sample sizes are normal to quite large for experimental studies of memory. The simulation conditions resulted from the fully crossed combination of the four design factors, resulting in 108 simulation conditions. For each of these, 1,000 response matrices were generated. For differences in dependency between conditions, the first experimental condition served as the reference condition. For the second experimental condition, the difference value was added to the dependency value of the first condition (i.e., the baseline dependency). Dependency values of zero indicate independence. For values larger zero there is positive dependency in the data. If the dependency difference is zero, the two experimental conditions are identical. Consequently, regarding results for testing against independence, only the results of the first experimental condition are reported. One limitation of D_{Q_3} is that the corresponding IRT model can not be estimated if there are items without variance because this prevents the estimation of item parameters for these items. To circumvent this problem in the simulation, the simulated data was redrawn until all items had non-zero variances.

The five dependency measures (D_{HB}, Q, Q_{a}[1], D_{Q_3}[2], and $D^{\text{np}}_{\text{Q}_3}$) were computed for each generated response matrix. Empirical detection rates were determined with the conventional significance level of $\alpha = 5\%$ using one-tailed testing[3] (dependency larger than zero for tests against independence and dependency lower in the first experimental condition than in the second experimental condition for tests of dependency differences). For D_{HB}, Q, and Q_{a} one-sample t-tests against zero were conducted for tests against independence and paired t-tests were conducted for tests of dependency differences. For the parametric bootstrap required for D_{Q_3}, the true parameter values (for fixed parameters) and correct distributional assumptions were used[4]. For each simulation condition, 1,000 bootstrap samples (cf. Davison & Hinkley, 1997) were generated prior to the simulation to obtain critical values for D_{Q_3}. Note that item and person parameters were only drawn once per simulation condition for the parametric bootstrap. For $D^{\text{np}}_{\text{Q}_3}$, 1,000 bootstrap samples were generated

---

[1] Q_{a} was computed by adding the constant 0.5 to each contingency table.

[2] While it may conceptually often make sense to set the guessing parameter to the stochastic guessing probability given some number of response alternatives, the true guessing parameters in the sample may deviate from this probability, for example due to participants using strategies that increase their probability of a correct response. Thus, we computed D_{Q_3} with different degrees of misspecification of the guessing parameter — $g = 0.2$ (no misspecification), $g = 0.15$ (underestimation), and $g = 0.25$ (overestimation). Over- or underestimation of the guessing parameter did not substantially affect the results and only the results with no misspecification of the guessing parameter are reported.

[3] We used one-tailed testing because the data generation process does not allow for negative dependencies (variances of the event-specific latent traits can not be negative).

[4] In practice one would have to estimate item parameters and latent trait variances from the data by initially fitting a unidimensional model (for tests against independence) or a bifactor model (for tests for differences between conditions).

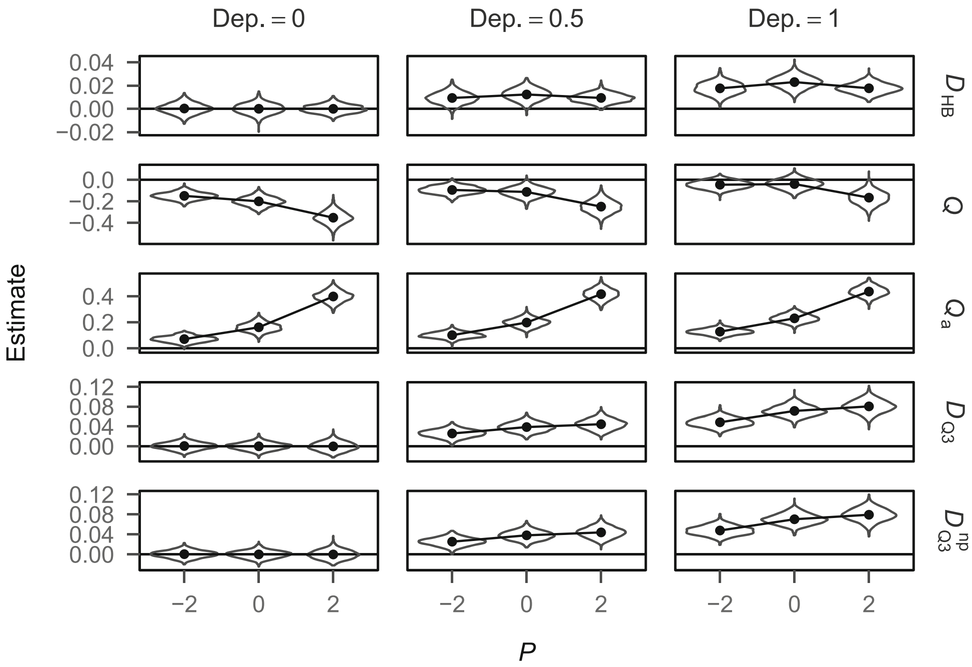


Fig. 1 Dependency estimates and mean trajectories obtained from the different measures by dependency and performance for $N = 100$. For D_{HB}, Q, and Q_{a} the displayed values refer to the mean across participants within the different simulation conditions. Note the varying y scales for the different measures.

from each generated response matrix. These were used for the nonparametric estimation of Q_3 (Debelak & Koller, 2020) and used to obtain critical values for $D_{\mathrm{Q}_3}^{\mathrm{np}}$.

The simulation was conducted in the R Programming Environment (R Core Team, 2021) using the packages *SimDesign* (version 2.2, Chalmers & Adkins, 2020), *mirt* (version 1.33.2, Chalmers, 2012), and *eRm* (version 1.0-1, Mair et al., 2020; Mair and Hatzinger, 2007)[5], and adapted functions from the package *sirt* (version 3.9-4, Robitzsch, 2020). Data and code for the simulation study are available via the Open Science Framework (OSF, https://osf.io/25mzu/).

Results

Figures referring to the distribution of dependency estimates (Figs. 1 and 4) show the values for a sample size of $N = 100$. Results for other sample sizes showed identical trends but distributions were more spread out due to larger standard errors. Because D_{HB}, Q, and Q_{a} yield participant-specific estimates, the values shown in the figures refer to the respective means across participants. This applies for both types of tests (i.e., tests against independence and tests for differences in dependency between experimental conditions).

Testing Against Independence

Estimates Figure 1 shows the distribution of dependency estimates yielded by the different approaches for the different simulation conditions. Given no dependency in the data, D_{HB}, D_{Q_3}, and $D_{\mathrm{Q}_3}^{\mathrm{np}}$ were distributed around zero across performance conditions. Q on the other hand was negatively biased and Q_{a} was positively biased and both biases increased strongly with performance. All estimates increased with increasing dependency in the data. The sensitivity of Q and Q_{a} to performance was maintained if there was dependency in the data. In such cases, D_{HB}, D_{Q_3}, and $D_{\mathrm{Q}_3}^{\mathrm{np}}$ also showed sensitivity to performance and this sensitivity increased with increasing dependency in the data, suggesting an interaction effect of dependency and performance on the estimates. D_{HB} showed the least sensitivity to performance and followed a curvilinear trend across performance conditions. D_{Q_3} and $D_{\mathrm{Q}_3}^{\mathrm{np}}$ showed similar sensitivity to performance with a monotonic increase in estimates across performance conditions. Sensitivity to performance was higher than for D_{HB} but was still very small compared to Q and Q_{a}.

In summary, D_{HB}, D_{Q_3}, and $D_{\mathrm{Q}_3}^{\mathrm{np}}$ were robust against different degrees of overall performance given that there was no dependency in the data but were sensitive to performance if there was dependency in the data. This sensitivity increased with increasing dependency and was less pronounced for mean values of D_{HB}. Q and Q_{a} were negatively and positively biased respectively and means

---

[5] The *eRm* package was used for computing $D_{\mathrm{Q}_3}^{\mathrm{np}}$. To do this, some of the package functions needed to be adjusted. The adjusted functions are available via the OSF.

were strongly affected by performance, even if there was no dependency in the data. Correlations between estimates of the different measures are shown in Table 1 in the Appendix.

Type I error rates Figure 2 shows the Type I error rates of the different approaches for the different simulation conditions. Q_a is not displayed because it yielded very high Type I error rates ($>$ .41), which strongly increased with performance. This can be explained by its positive bias (see Fig. 1) and the one-tailed testing applied. Q is also not displayed because it yielded Type I error rates of zero in all conditions, which can be explained by its negative bias (see Fig. 1) and the one-tailed testing applied. D_{HB} tended to yield higher Type I error rates than D_{Q_3} and $D_{\mathrm{Q}_3}^{\mathrm{np}}$ except for smaller sample sizes. There was no clear trend of Type I error rates across performance conditions, suggesting that the three measures yield Type I error rates that are unaffected by performance. D_{Q_3} and $D_{\mathrm{Q}_3}^{\mathrm{np}}$ yielded Type I error rates close to 5%, suggesting good maintenance of the nominal significance level by these measures.

Power Figure 3 shows the power of the different approaches for detecting dependency for the different simulation conditions. Power increased with sample size and increasing dependency in the data. Q yielded very low power, which can again be explained by its negative bias (see Fig. 1). Q_a yielded very high power that is sensitive to performance. This can be explained by the measure's positive bias (see Fig. 1). D_{Q_3} and $D_{\mathrm{Q}_3}^{\mathrm{np}}$ yielded comparable power that was higher than the one yielded by D_{HB}. The power yielded by all three measures was sensitive to performance but this sensitivity was comparable between the three measures.

Testing for differences in dependency

Estimates Figure 4 shows the distribution of estimates of dependency differences yielded by the different approaches for the different simulation conditions. Given no difference between conditions, all estimates were distributed around zero, irrespective of performance and baseline dependency (i.e., dependency in the reference condition). All estimates decreased with increasing differences in dependency in the data. If there were dependency differences in the data, D_{HB} showed the least sensitivity to performance and followed a curvilinear trend across performance conditions. Q and Q_a were highly sensitive to performance. While Q monotonically increased with increasing performance, Q_a followed a curvilinear trend across performance conditions. D_{Q_3} and $D_{\mathrm{Q}_3}^{\mathrm{np}}$ showed similar sensitivity to performance with a monotonic decrease in estimates across performance conditions. Sensitivity to performance was higher than for D_{HB} but was smaller than for Q and Q_a. Sensitivity to memory performance increased with increasing differences in dependency for all measures. Finally, all estimates shifted closer to zero with an increasing baseline dependency. Correlations between estimates of dependency differences of the different measures are shown in Table 2 in the Appendix.

Type I error rates Figure 5 shows the Type I error rates of the different approaches when testing for differences in

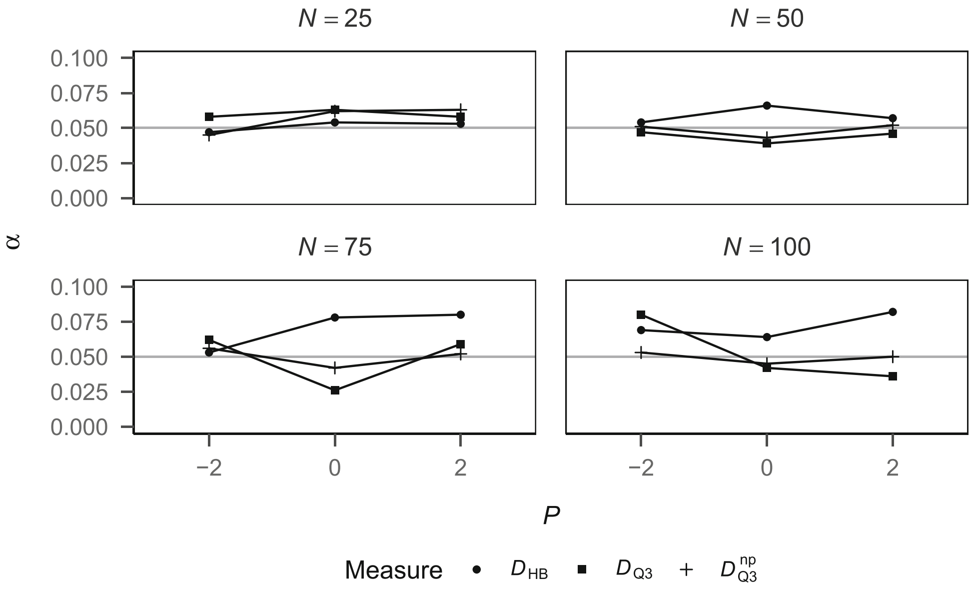


Fig. 2 Type I error rates of the different measures for tests against independence by performance and sample size. Q and Q_a are not displayed.

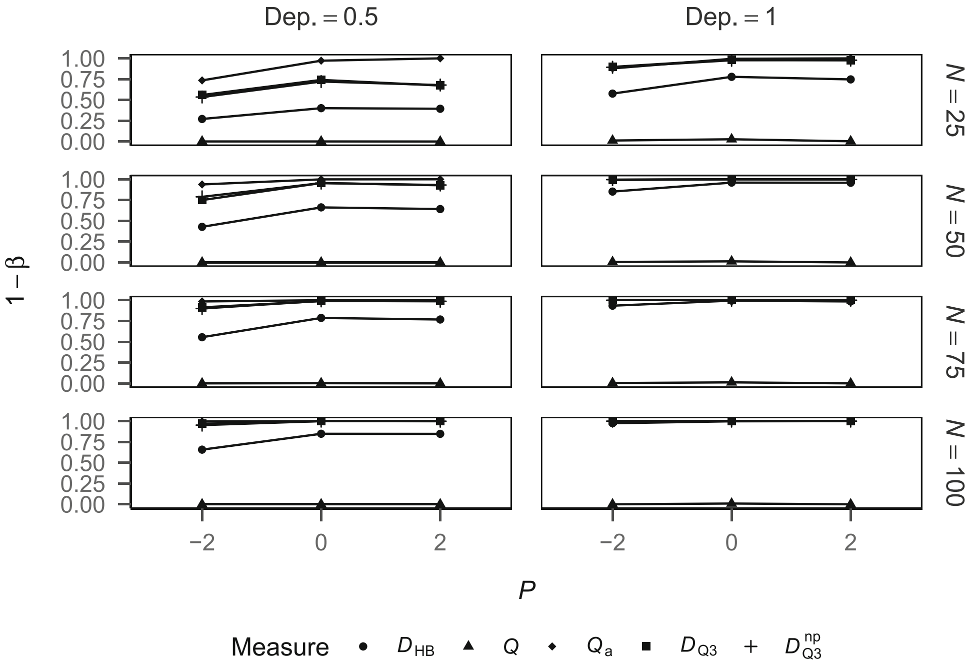


Fig. 3 Power of the different measures for detecting dependency by performance, baseline dependency, and sample size

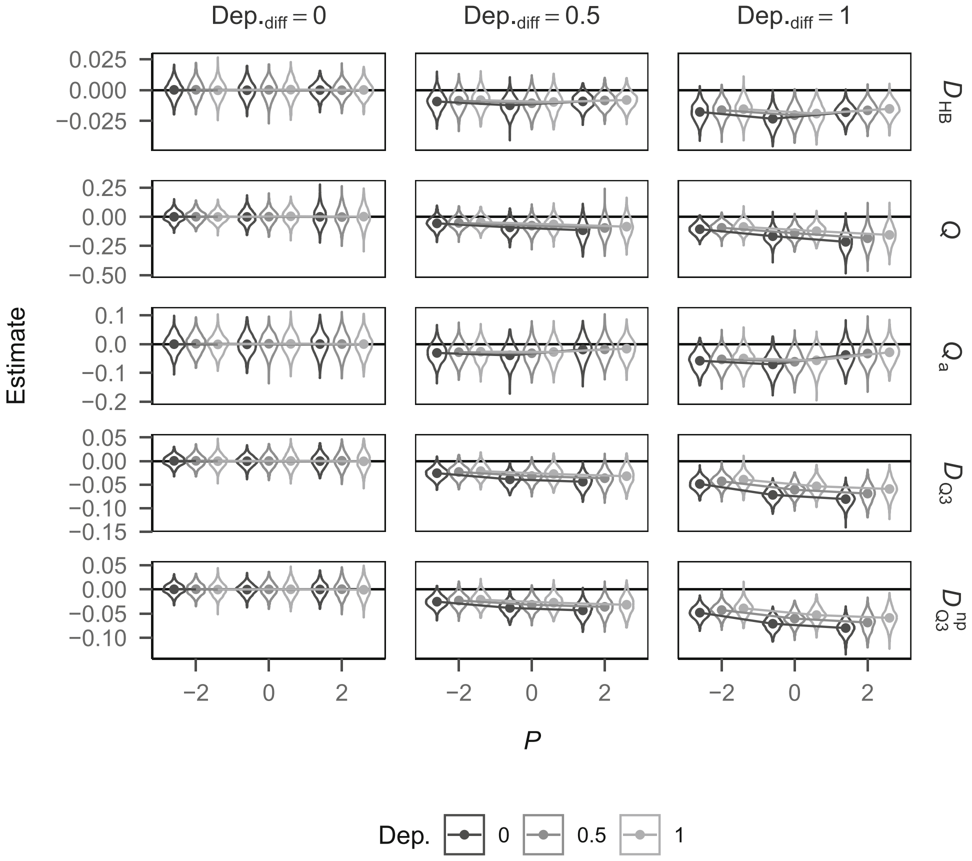


Fig. 4 Estimates and mean trajectories of dependency differences obtained from the different measures by baseline dependency, dependency difference, and performance for $N = 100$. For D_{HB}, Q, and Q_a the displayed values refer to the mean differences across participants within the different simulation conditions. Note the varying y scales for the different measures.

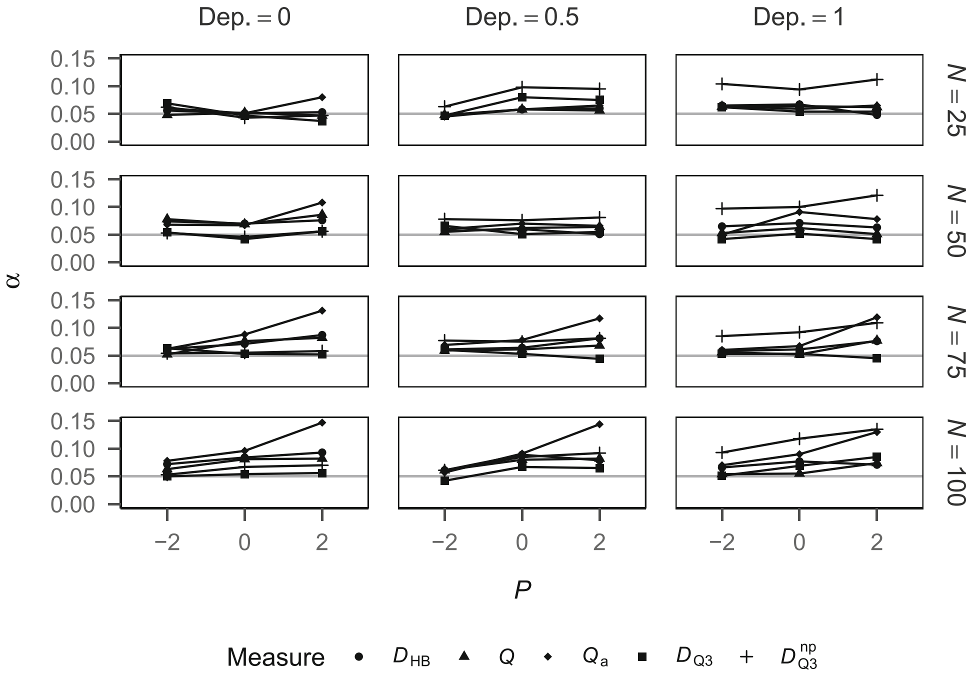


Fig. 5 Type I error rates of the different measures for tests for differences in dependency by performance, baseline dependency, and sample size

dependency for the different simulation conditions. Q_{a} and $D_{\mathrm{Q}_3}^{\mathrm{np}}$ yielded the highest Type I error rates, whereas Type I error rates for D_{HB}, Q, and D_{Q_3} were approximately comparable. Overall, D_{Q_3} showed the best maintenance of the nominal significance level. For $D_{\mathrm{Q}_3}^{\mathrm{np}}$, Type I error rates increased with increasing baseline dependency. This was not the case for the other measures. There was no clear trend of Type I error rates across performance conditions, suggesting that the Type I error rates of the measures are unaffected by performance, except for Q_{a} for which Type I error rates increased with performance for larger sample sizes.

Power Figure 6 shows the power of the different approaches for detecting differences in dependency for the different simulation conditions. Power increased with sample size and increasing dependency differences in the data and decreased with increasing baseline dependency for all measures. Q_{a} yielded the lowest power, followed by Q, and both measures were highly sensitive to performance, with a curvilinear trend across performance conditions. D_{HB} yielded higher power than Q_{a} and Q but lower power than D_{Q_3} and $D_{\mathrm{Q}_3}^{\mathrm{np}}$. D_{HB} was sensitive to performance, either monotonically increasing with performance or showing a curvilinear trend across performance conditions, but the sensitivity to performance was lower than for Q_{a} and Q. D_{Q_3} and $D_{\mathrm{Q}_3}^{\mathrm{np}}$ yielded the highest power, with slightly higher power for $D_{\mathrm{Q}_3}^{\mathrm{np}}$ than D_{Q_3}. This difference increased with increasing baseline dependency and may be explained by the increased sensitivity of $D_{\mathrm{Q}_3}^{\mathrm{np}}$ given a higher level of dependency in the data, which also manifested in higher Type I error rates (see Fig. 5). D_{Q_3} and $D_{\mathrm{Q}_3}^{\mathrm{np}}$ were similarly sensitive to performance as D_{HB}, either monotonically increasing with performance or showing a curvilinear trend across performance conditions.

Discussion

The simulation showed that Q yields negatively biased and Q_{a} yields positively biased estimates, even if there is no dependency in the data. This also manifests in very high Type I error rates for Q_{a} and very low power for detecting stochastic dependency of the retrieval of event elements for Q. The measures perform somewhat better when testing for differences in dependency between experimental conditions but are still inferior to the other measures. The two measures are also strongly affected by varying levels of overall memory performance, since they do not attempt to correct for memory performance as do D_{HB}, D_{Q_3}, and $D_{\mathrm{Q}_3}^{\mathrm{np}}$. The latter three measures yield unbiased estimates and are unaffected by varying levels of overall memory performance given no dependency or no difference in dependency. However, if there is dependency or there are differences in dependency, all three measures are affected by memory performance, although to a much smaller extent than Q and Q_{a}. In such cases, the power of D_{HB}, D_{Q_3}, and $D_{\mathrm{Q}_3}^{\mathrm{np}}$ is affected to a similar degree, even though the mean estimates of

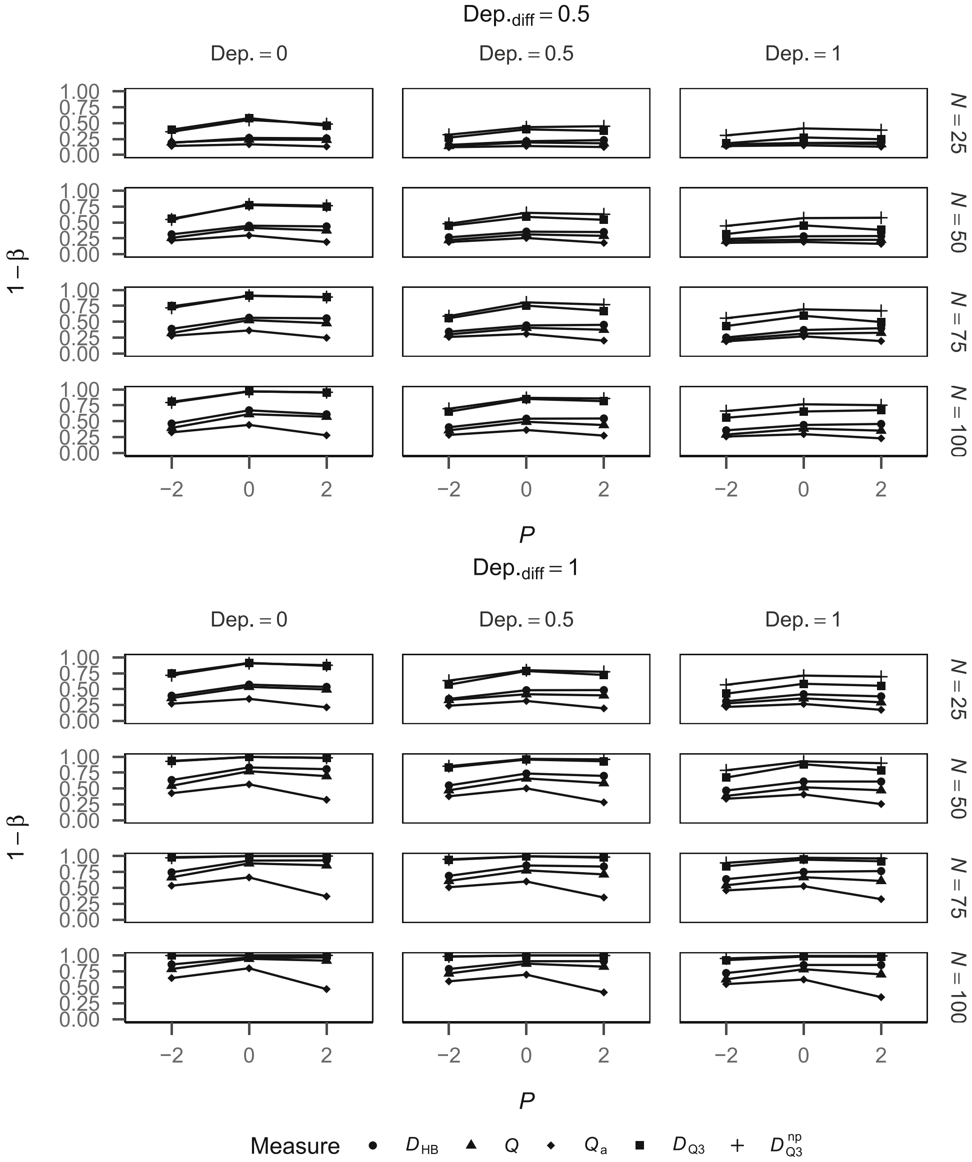


Fig. 6 Power of the different measures for detecting differences in dependency by performance, baseline dependency, sample size, and dependency difference

D_{HB} across participants are least affected by memory performance. Note however, that person-specific estimates may be more strongly affected by memory performance. D_{HB} is affected by memory performance because the data-based dependency estimate and the dependency estimate from the independent model do not scale perfectly equal with memory performance. For D_{Q_3} this may be because fitting a unidimensional IRT model to locally dependent data leads to overestimation of measurement precision (Ip, 2010; Wainer and Wang, 2000) and worse recovery of person parameters (Koziol, 2016). Similar problems may arise for $D_{\mathrm{Q}_3}^{\mathrm{np}}$. While it does not require the estimation of person parameters, it builds on the property of sum scores as sufficient statistics in the Rasch model (Rasch, 1960), which assumes local independence. D_{Q_3} and $D_{\mathrm{Q}_3}^{\mathrm{np}}$ yield higher power than D_{HB}, emphasizing the advantage of running analyses on individual item responses rather than aggregated contingency tables. However, when testing for differences in dependency, $D_{\mathrm{Q}_3}^{\mathrm{np}}$ yields increased Type I error rates with increasing dependency in the data. Since D_{HB} and D_{Q_3} are unbiased under the null hypothesis and their Type I error rates are unaffected by memory performance and baseline dependency (for $D_{\mathrm{Q}_3}^{\mathrm{np}}$ this holds for single parameter tests, but not for tests of parameter differences), their susceptibility to memory performance reduces to a power problem when focusing on statistical inferences rather than descriptive estimates.

Overall, D_{Q_3} performed best because it yields unbiased estimates under the null hypothesis, provides good maintenance of Type I error rates that tend to be better than that of D_{HB} and $D^{\mathrm{np}}_{\mathrm{Q}_3}$, especially when testing for differences in dependency, and yields high power, although power is similarly affected by memory performance as is the power of D_{HB} and $D^{\mathrm{np}}_{\mathrm{Q}_3}$. Next, we applied the different measures to an empirical example to compare the congruence of inferences drawn from empirical data.

Empirical application

Methods

As an empirical data example, a dataset by James et al. (2020, Experiment 1), was used (the original data is available at https://osf.io/cqm7v/). In this experiment 45 participants were presented events consisting of an animal, an object, and a location. Event elements were presented as cartoon illustrations, which were additionally named aloud through headphones. There were 2 experimental conditions, which were administered in a within-subjects design and with 15 events presented in each condition. In the simultaneous encoding condition all event elements were presented together in a single learning trial. In the separated encoding condition each pairwise association between event elements was presented separately across three learning trials. After encoding, participants conducted a cued recognition test with four response alternatives and six test trials per event (all associations were tested in both directions), resulting in 180 items. Mean memory performance was .71 in the simultaneous encoding condition and .73 in the separated encoding condition, making the setting similar to the simulation conditions with P = 2. Previous studies found a significant positive dependency in both a simultaneous and a separated encoding condition, with no significant difference in dependency between conditions (Bisby et al., 2018; Horner and Burgess, 2014).

The five dependency measures were computed based on the data, using a significance level of $\alpha = 5\%$ (two-tailed testing). For computing D_{Q_3}, g was set to the stochastic guessing probability of 0.25 given four response alternatives[6]. The analysis scripts for the empirical application are available via the OSF (https://osf.io/25mzu/).

Results

Results using the different dependency measures are shown in Fig. 7. The results for D_{HB} are in accordance with those reported by James et al. (2020) — there was a significant positive dependency in the simultaneous encoding condition but not in the separated encoding condition, with a significant difference between conditions. This contradicts previous findings by Horner and Burgess (2014) and Bisby et al. (2018), which found a significant positive dependency also in the separated encoding condition and no difference in dependency between conditions. D_{Q_3} and $D^{\mathrm{np}}_{\mathrm{Q}_3}$ yielded similar results as D_{HB}. However, using these measures the dependency in the separated encoding condition was also positive and significant, with the difference between conditions still being significant. Since D_{Q_3} and $D^{\mathrm{np}}_{\mathrm{Q}_3}$ yield higher power for detecting dependencies than D_{HB} it may be the case that the power of D_{HB} is insufficient for detecting the weak dependency in the separated encoding condition. The results using D_{Q_3} and $D^{\mathrm{np}}_{\mathrm{Q}_3}$ are also more consistent with the findings by Horner and Burgess (2014) and Bisby et al. (2018) in the sense that they also found a positive dependency in the separated encoding condition. However, they are consistent with the finding by James et al. (2020) that there is a significant difference in dependency between conditions, which was not found by Horner and Burgess (2014) and Bisby et al. (2018).

Q and Q_{a} yielded very different results than the other measures. Using Q, there was no significant dependency in the simultaneous encoding condition and a significantly negative dependency in the separated encoding condition, with a significant difference between conditions. Using Q_{a} there was a significant positive dependency in both conditions but the difference between conditions was nonsignificant. These divergent findings may be explained by the negative bias of Q and the positive bias of Q_{a}. The results using Q are quite inconsistent with previous findings and are only partially consistent with the findings by James et al. (2020) in the sense that there is a significant difference in dependency between conditions. While the results using Q_{a} are actually in accordance with the findings by Horner and Burgess (2014) and Bisby et al. (2018), the results from the simulation study and the incongruence with results using the other measures indicate that this result is likely not a correct representation of the given data.

General discussion

In the current research we compared five approaches for measuring binding effects (i.e., stochastic dependencies of the retrieval of event elements) in event-based episodic representations regarding their empirical detection rates,

---

[6] Given that associations were tested in both directions, it may be possible that guessing differed between the first and second test of an association within an event. However, in the absence of more specific information and also considering model parsimony, we considered the stochastic guessing probability to be the most objective and appropriate criterion.

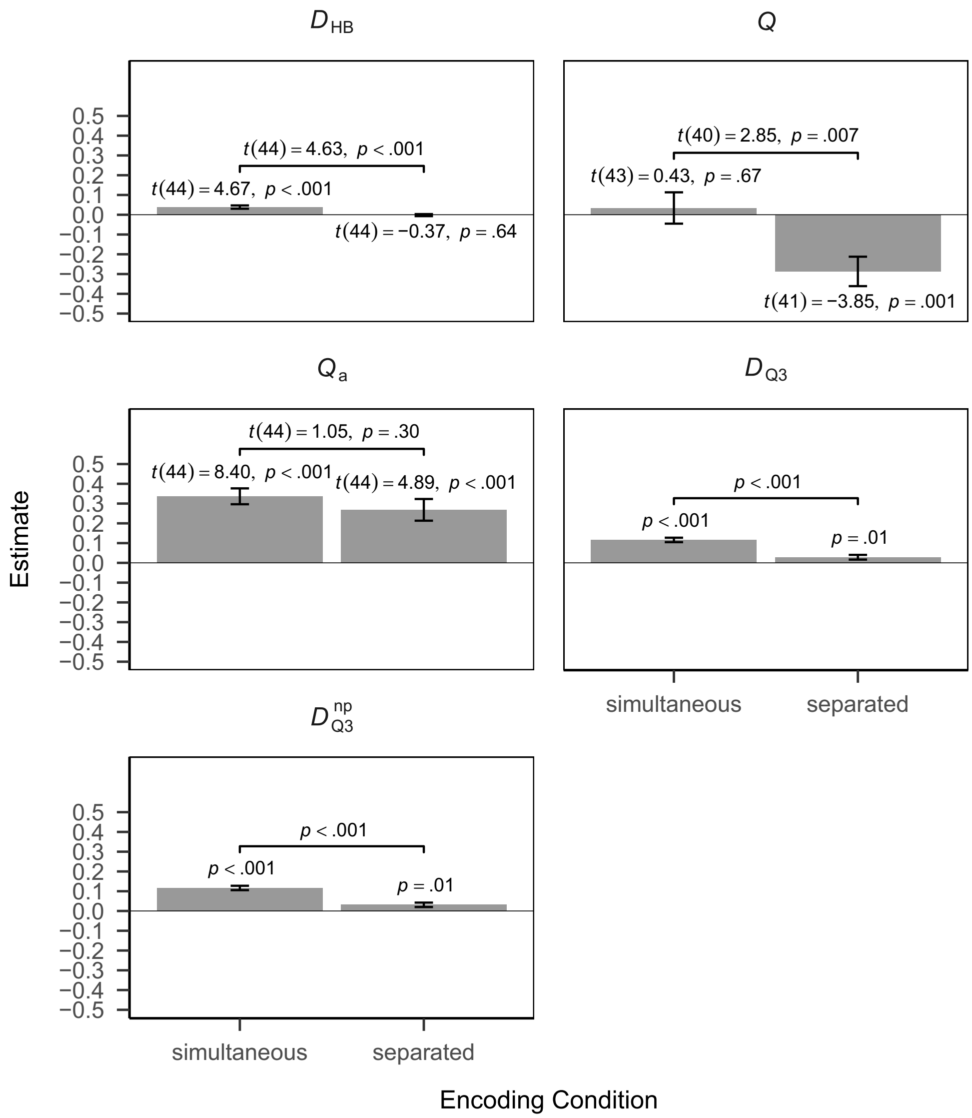


Fig. 7 Results for the data of Experiment 1 by James et al. (2020) using the different approaches

susceptibility to memory performance, and congruence of empirical estimates. The approaches based on Yule's Q (Q and Q_a, Yule, 1912; cf. Horner & Burgess, 2014) yield biased estimates, with Q being negatively and Q_a being positively biased. In addition, the measures are highly susceptible to memory performance and applying them to the empirical example lead to considerable deviations from the results obtained by applying the other approaches. Thus, Q and Q_a are unsuitable for measuring binding effects in event-based episodic representations. The approach by Horner and Burgess (2013, D_{HB}), the IRT-based approach (D_{Q_3}, Schreiner et al., in press), and the nonparametric variant of the IRT-based approach ($D_{Q_3}^{np}$, cf. Debelak and Koller, 2020; Schreiner et al., in press) are unbiased and not susceptible to memory performance under the null hypothesis of no dependency in the data or no differences in dependency between conditions. They are however affected by performance if there is dependency in the data or there are differences in dependency between conditions. This affects the power of all three measures to a similar degree. Since memory performance affects the power but not the Type I error rates of these measures, they do not elicit artifactual binding results as a consequence of base performance. This is because, when focusing on statistical inferences, the sensitivity of the measures is only affected if there is a true binding effect, reducing the effect of memory performance to a power problem. D_{Q_3} and $D_{Q_3}^{np}$ yield higher power than D_{HB}. However, $D_{Q_3}^{np}$ yields increased Type I error rates with increasing dependency in the data when testing for differences in dependency between conditions. Compared to D_{Q_3}, $D_{Q_3}^{np}$ yielded, on average, Type I error rates increased by 0.003

if the baseline dependency was 0, 0.02 if the baseline dependency was 0.5, and 0.05 if the baseline dependency was 1. Applying D_{HB}, D_{Q_3}, and $D_{\mathrm{Q}_3}^{\mathrm{np}}$ to the empirical example lead to similar results, but the results obtained by applying D_{Q_3} and $D_{\mathrm{Q}_3}^{\mathrm{np}}$ were more consistent with previous findings by Horner and Burgess (2014) and Bisby et al. (2018). Given that memory performance in the empirical example was relatively high and similar to the simulation conditions with $P = 2$, the estimates for D_{Q_3} and $D_{\mathrm{Q}_3}^{\mathrm{np}}$ may be somewhat inflated, given that estimates for these measures tend to increase with performance, and more so than the mean values of D_{HB}. However, Type I error rates of the two measures do not increase with performance. Thus, the statistical inference that there is a significant positive dependency in the separated encoding condition can not be attributed to inflated sensitivity of D_{Q_3} and $D_{\mathrm{Q}_3}^{\mathrm{np}}$ due to high memory performance. Taking together the simulation results and the results from the empirical application, D_{Q_3} performed best among the five measures. It provides unbiased estimates under the null hypothesis, provides good maintenance of Type I error rates that are unaffected by memory performance and baseline dependency, yields high power (subject to memory performance like D_{HB} and $D_{\mathrm{Q}_3}^{\mathrm{np}}$), and yielded results for the empirical example that are more consistent with findings of previous studies (Bisby et al., 2018; Horner and Burgess, 2014).

A potential limitation concerning the results may be that both D_{Q_3} and the data generation procedure were IRT-based and we used the true discrimination parameters and distributional assumptions for computing D_{Q_3}. This may have provided D_{Q_3} with some advantage over the other approaches. However, we chose the data generation procedure because it reflects well the actual psychological processes in memory retrieval given binding effects. In that sense, one could argue that D_{Q_3} is a better approximation of the psychological processes that underlie binding effects than are the other approaches. Further, D_{Q_3} should be rather robust against misspecifications of certain model parameters or distributional assumptions, since such misspecifications affect both within- and between-event residual correlations, which are contrasted in the computation of D_{Q_3}. The finding that misspecification of the guessing parameter in the simulation study did not substantially affect the results supports this notion. Nevertheless, the robustness of D_{Q_3} against misspecifications of model parameters and distributional assumptions should be examined in future research.

D_{Q_3} provides some additional advantages. First, it operates on the level of individual item responses rather than aggregate contingency tables as do D_{HB}, Q, and Q_{a}, and the IRT model on which the measure is based considers participant and item differences as well as participant-item interactions. Thus, contrary to the contingency-based approaches, D_{Q_3} is not prone to Simpson's paradox (Hintzman, 1972, 1980; Simpson, 1951; see also Burton et al., 2017). Second, IRT-based measures enable established and plausible modeling of meaningful psychological variables instead of running analyses on the basis of descriptive contingency tables. Third, D_{Q_3} can in principle be applied to a greater variety of testing procedures than the contingency-based approaches. The contingency-based approaches require some common feature of items for identifying the dependency pairs, such as items having a common cue or target element. If testing situations do not involve cueing, such identifying features are absent and consequently, dependency pairs would be arbitrary. Since D_{Q_3} does not require such identifying features (the assignment of items to a common event is sufficient), it can in principle also be applied to testing situations not involving cueing such as free recall or free recognition. For example, imagine participants are presented three words in a joint temporal context at a time, forming an event. Then, each word can form a binary item that is assigned the value 1, if the word has been successfully recalled or recognized, and 0 if not, resulting in three items per event. One can then compute D_{Q_3} the same way as for cued recognition output, based on the residual correlations between item pairs. Yet, evaluating the consistency of D_{Q_3}, and also the other approaches, across different types of memory tests is an interesting prospect for future research. This would likely require a systematic investigation of several empirical data sets, which used various types of memory tests, or conducting an experiment with a given paradigm and varying the type of memory test between participants. Fourth, D_{Q_3} can in principle be extended to account for polytomous instead of dichotomous item responses, for example by using the rating scale model (Andrich, 1978) or the partical credit model (Masters, 1982) as the basis for computing the Q_3 statistics. Finally, the approach yields estimated person and item parameters as useful by-products of the dependency analysis. For example, in applications with fixed event composition rather than random assignment of elements to events, item parameters may be used to identify problematic events with, for example, very high or very low difficulty of the associated items to improve the study material for subsequent experiments. Person parameters may be used to compare participants regarding their overall memory performance (but note that estimation of person parameter may be negatively affected by binding effects resulting in locally dependent data, see Koziol, 2016). However, some further considerations have to be taken into account when selecting a suitable measure for a given setting.

First, D_{Q_3} yields an overall or condition-specific dependency estimate. In some cases it may be necessary to obtain person-specific dependency estimates, which are not provided by D_{Q_3} in its current implementation. These are however provided by D_{HB} and one may use this measure in such cases. Second, if one wants to use D_{Q_3} and there are items without variance, item parameters for these items can not be estimated. In such a case one would have to exclude these items or reorder items if possible. The risk of this to occur increases with smaller sample sizes, increasing prevalence of missing values, and more extreme levels of memory performance. In the simulation, this issue was actively prevented by resampling until there were no items without variance. Still, there were some convergence issues for small sample sizes. Third, the bootstrap approach for D_{Q_3} is currently only designed for the comparison of two conditions, thus only enabling pairwise comparisons when using D_{Q_3}. Finally, power is not the only issue to consider when determining sample size when using D_{Q_3}. Parameter estimation becomes more stable with increasing sample size. This leads to more reliable estimates and may enable one to freely estimate parameters that may have to be fixed for smaller sample sizes, for example discrimination or guessing parameters, making the measure more flexible. In summary, we recommend to use D_{Q_3} as a measure of binding effects in event-based episodic representations if the mentioned considerations have been taken into account.

Acknowledgements We thank Alicia Gernand for helping us test the simulation code.

Funding Open Access funding enabled and organized by Projekt DEAL. This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - GRK 2277 “Statistical Modeling in Psychology.”

Declarations

Data and Code Availability The data and code are available via the Open Science Framework (https://osf.io/25mzu/). The study was not preregistered.

Ethics Approval Not applicable.

Consent to participate Not applicable.

Consent to publish Not applicable.

Conflict of Interests We have no relevant financial or non-financial interests to disclose.

Appendix: Correlations Between Dependency Estimates

Table 1 Mean [Range] of Correlations Between Dependency Estimates of the Different Measures Across Simulation Conditions

	D_{HB}	Q	Q_{a}	D_{Q_3}
D_{HB}	1			
Q	.82 [.54, .94]	1		
Q_{a}	.65 [.13, .93]	.41 [-.19, .85]	1	
D_{Q_3}	.78 [.44, .94]	.64 [.23, .89]	.50 [.12, .83]	1
$D_{\mathrm{Q}_3}^{\mathrm{np}}$	.79 [.44, .94]	.64 [.24, .89]	.50 [.13, .84]	.99 [.96, $>$ .99]

Notes. For D_{HB}, Q, and Q_{a} correlations refer to the mean values of the respective estimates. For computing the mean correlations, Fisher’s Z-transformation was applied.

Table 2 Mean [Range] of Correlations Between Estimates of Dependency Differences of the Different Measures Across Simulation Conditions

	D_{HB}	Q	Q_{a}	D_{Q_3}
D_{HB}	1			
Q	.78 [.65, .85]	1		
Q_{a}	.85 [.60, .93]	.64 [.19, .83]	1	
D_{Q_3}	.60 [.41, .70]	.44 [.31, .58]	.50 [.25, .66]	1
$D_{\mathrm{Q}_3}^{\mathrm{np}}$	.61 [.44, .71]	.45 [.32, .58]	.51 [.27, .67]	.98 [.96, .99]

Notes. For D_{HB}, Q, and Q_{a} correlations refer to the mean values of the respective difference estimates. For computing the mean correlations, Fisher’s Z-transformation was applied.

References

Andermane, N., Joensen, B. H., & Horner, A. J. (2021). Forgetting across a hierarchy of episodic representations. *Current Opinion* *in Neurobiology*, *67*, 50–57. https://doi.org/10.1016/j.conb.2020.08.004

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, *43*(4), 561–573. https://doi.org/10.1007/bf02293814

Arnold, N. R., Heck, D. W., Bröder, A., Meiser, T., & Boywitt, C. D. (2019). Testing hypotheses about binding in context memory with a hierarchical multinomial modeling approach: A preregistered study. *Experimental Psychology*, *66*(3), 239–251. https://doi.org/10.1027/1618-3169/a000442

Balaban, H., Assaf, D., Arad Meir, M., & Luria, R. (2019). Different features of real-world objects are represented in a dependent manner in long-term memory. *Journal of Experimental Psychology: General, 149*(7). https://doi.org/10.1037/xge0000716

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In Lord, F. M., & Novick, M. R. (Eds.) *Statistical theories of mental test scores*: Addison-Wesley.

Bisby, J. A., Horner, A. J., Bush, D., & Burgess, N. (2018). Negative emotional content disrupts the coherence of episodic memories. *Journal of Experimental Psychology: General*, *147*(2), 243–256. https://doi.org/10.1037/xge0000356

Boywitt, C. D., & Meiser, T. (2012a). Bound context features are integrated at encoding. *Quarterly Journal of Experimental Psychology*, *65*(8), 1484–1501. https://doi.org/10.1080/17470218.2012.656668

Boywitt, C. D., & Meiser, T. (2012b). The role of attention for context-context binding of intrinsic and extrinsic features. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(4), 1099–1107. https://doi.org/10.1037/a0026988

Burton, R. L., Lek, I., & Caplan, J. B. (2017). Associative independence revisited: Competition between conflicting associations can be resolved or even reversed in one trial. *Quarterly Journal of Experimental Psychology*, *70*(4), 832–857. https://doi.org/10.1080/17470218.2016.1171886

Burton, R. L., Lek, I., Dixon, R. A., & Caplan, J. B. (2019). Associative interference in older and younger adults. *Psychology and Aging*, *34*(4), 558–571. https://doi.org/10.1037/pag0000361

Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, *48*(6), 1–29. https://doi.org/10.18637/jss.v048.i06

Chalmers, R. P., & Adkins, M. C. (2020). Writing effective and reliable Monte Carlo simulations with the SimDesign package. *The Quantitative Methods for Psychology*, *16*(4), 248–280. https://doi.org/10.20982/tqmp.16.4.p248

Chen, W.-H., & Thissen, D. (1997). Local dependence indexes for item pairs using item response theory. *Journal of Educational and Behavioral Statistics*, *22*(3), 265–289. https://doi.org/10.3102/10769986022003265

Cohen, J. (1988). Statistical power analysis for the behavioral sciences, 2nd ed. Erlbaum.

Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their application*. Cambridge University Press.

de Ayala, R. J. (2009). *The theory and practice of item response theory*. Guilford Press.

Debelak, R., & Koller, I. (2020). Testing the local independence assumption of the rasch model with Q3,-based nonparametric model tests. *Applied Psychological Measurement*, *44*(2), 103–117. https://doi.org/10.1177/0146621619835501

Gibbons, R. D., & Hedeker, D. R. (1992). Full-information item bi-factor analysis. *Psychometrika*, *57*(3), 423–436. https://doi.org/10.1007/bf02295430

Goodman, L. A., & Kruskal, W. H. (1954). Measures of association for cross classifications. *Journal of the American Statistical Association*, *49*(268), 732–764. https://doi.org/10.1080/01621459.1954.10501231

Hayman, C. G., & Tulving, E. (1989). Contingent dissociation between recognition and fragment completion: The method of triangulation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(2), 228–240. https://doi.org/10.1037/0278-7393.15.2.228

Hicks, J. L., & Starns, J. J. (2016). Successful cuing of gender source memory does not improve location source memory. *Memory & Cognition*, *44*(4), 650–659. https://doi.org/10.3758/s13421-016-0586-y

Hintzman, D. L. (1972). On testing the independence of associations. *Psychological Review*, *79*(3), 261–264. https://doi.org/10.1037/h0032684

Hintzman, D. L. (1980). Simpson's paradox and the analysis of memory retrieval 87(4), 398–410. https://doi.org/10.1037/0033-295x.87.4.398

Horner, A. J., Bisby, J. A., Bush, D., Lin, W.-J., & Burgess, N. (2015). Evidence for holistic episodic recollection via hippocampal pattern completion. *Nature Communications*, *6*(1), 7462. https://doi.org/10.1038/ncomms8462

Horner, A. J., & Burgess, N. (2013). The associative structure of memory for multi-element events. *Journal of Experimental Psychology: General*, *142*(4), 1370–1383. https://doi.org/10.1037/a0033626

Horner, A. J., & Burgess, N. (2014). Pattern completion in multielement event engrams. *Current Biology*, *24*(9), 988–992. https://doi.org/10.1016/j.cub.2014.03.012

Hunt, R. R., & Einstein, G. O. (1981). Relational and item-specific information in memory. *Journal of Verbal Learning and Verbal Behavior*, *20*(5), 497–514. https://doi.org/10.1016/S0022-5371(81)90138-9

Ip, E. H. (2010). Interpretation of the three-parameter testlet response model and information function. *Applied Psychological Measurement*, *34*(7), 467–482. https://doi.org/10.1177/0146621610364975

James, E., Ong, G., Henderson, L., & Horner, A. J. (2020). Make or break it: Boundary conditions for integrating multiple elements in episodic memory. *Royal Society Open Science*, *7*(9), 200431. https://doi.org/10.1098/rsos.200431

Joensen, B. H., Gaskell, M. G., & Horner, A. J. (2020). United we fall: All-or-none forgetting of complex episodic events. *Journal of Experimental Psychology: General*, *149*(2), 230–248. https://doi.org/10.1037/xge0000648

Kahana, M. J. (2002). Associative symmetry and memory theory. *Memory & Cognition*, *30*(6), 823–840. https://doi.org/10.3758/BF03195769

Kahana, M. J., Rizzuto, D. S., & Schneider, A. R. (2005). Theoretical correlations and measured correlations: Relating recognition and recall in four distributed memory models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(5), 933–953. https://doi.org/10.1037/0278-7393.31.5.933

Koziol, N. A. (2016). Parameter recovery and classification accuracy under conditions of testlet dependency: a comparison of the traditional 2PL, testlet, and bi-factor models. *Applied Measurement in Education*, *29*(3), 184–195. https://doi.org/10.1080/08957347.2016.1171767

Lazarsfeld, P. F., & Henry, N. W. (1968). *Latent structure analysis*. Houghton Mifflin.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Erlbaum.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Addison-Wesley.

Mair, P., & Hatzinger, R. (2007). Extended Rasch modeling: The eRm package for the application of IRT models in R. *Journal of Statistical Software*, *20*(9), 1–20. https://doi.org/10.18637/jss.v020.i09

Mair, P., Hatzinger, R., & Maier, M. J. (2020). eRm: Extended Rasch Modeling. 1.0-1, https://cran.r-project.org/package=eRm.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, *47*(2), 149–174. https://doi.org/10.1007/bf02296272

Meiser, T., & Bröder, A. (2002). Memory for multidimensional source information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(1), 116–137. https://doi.org/10.1037//0278-7393.28.1.116

Ngo, C. T., Horner, A. J., Newcombe, N. S., & Olson, I. R. (2019). Development of holistic episodic recollection. *Psychological Science*, *30*(12), 1696–1706. https://doi.org/10.1177/0956797619879441

Ponocny, I. (2001). Nonparametric goodness-of-fit tests for the rasch model. *Psychometrika*, *66*(3), 437–459. https://doi.org/10.1007/BF02294444

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing. https://www.R-project.org/.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Nielsen & Lydiche.

Robin, J. (2018). Spatial scaffold effects in event memory and imagination. *WIREs Cognitive Science*, *9*(4), e1462. https://doi.org/10.1002/wcs.1462

Robitzsch, A. (2020). sirt: Supplementary item response theory models. *R package version, 3.9–4*. https://cran.r-project.org/package=sirt.

Rubin, D. C., & Umanath, S. (2015). Event memory: A theory of memory for laboratory, autobiographical, and fictional events. *Psychological Review*, *122*(1), 1–23. https://doi.org/10.1037/a0037907

Schreiner, M. R., Meiser, T., & Bröder, A. (in press). The binding structure of event elements in episodic memory and the role of animacy. *Quarterly Journal of Experimental Psychology*.

Simpson, E. H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society*, *13*(2), 238–241. https://doi.org/10.1111/j.2517-6161.1951.tb00088.x

Starns, J. J., & Hicks, J. L. (2005). Source dimensions are retrieved independently in multidimensional monitoring tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(6), 1213–1220. https://doi.org/10.1037/0278-7393.31.6.1213

Starns, J. J., & Hicks, J. L. (2008). Context attributes in memory are bound to item information, but not to one another. *Psychonomic Bulletin & Review*, *15*(2), 309–314. https://doi.org/10.3758/PBR.15.2.309

Utochkin, I. S., & Brady, T. F. (2020). Independent storage of different features of real-world objects in long-term memory. *Journal of Experimental Psychology: General*, *149*(3), 530–549. https://doi.org/10.1037/xge0000664

Verhelst, N. D. (2008). An efficient MCMC algorithm to sample binary matrices with fixed marginals. *Psychometrika*, *73*(4), 705–728. https://doi.org/10.1007/s11336-008-9062-3

Vogt, V., & Bröder, A. (2007). Independent retrieval of source dimensions: An extension of results by Starns, and Hicks (2005) and a comment on the ACSIM measure. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(2), 443–450. https://doi.org/10.1037/0278-7393.33.2.443

Wainer, H., & Wang, X. (2000). Using a new statistical model for testlets to score TOEFL. *Journal of Educational Measurement*, *37*(3), 203–220. https://doi.org/10.1111/j.1745-3984.2000.tb01083.x

Wang, W.-C., & Wilson, M. (2005). The Rasch, testlet model. *Applied Psychological Measurement*, *29*(2), 126–149. https://doi.org/10.1177/0146621604271053

Yen, W. M. (1984). Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Applied Psychological Measurement*, *8*(2), 125–145. https://doi.org/10.1177/014662168400800201

Yen, W. M. (1993). Scaling performance assessments: Strategies for managing local item dependence. *Journal of Educational Measurement*, *30*(3), 187–213. https://doi.org/10.1111/j.1745-3984.1993.tb00423.x

Yule, G. U. (1912). On the methods of measuring association between two attributes. *Journal of the Royal Statistical Society*, *75*(6), 579–652. https://doi.org/10.2307/2340126

Original Article

QJEP

The binding structure of event elements in episodic memory and the role of animacy

Quarterly Journal of Experimental Psychology
1–26

DOI: 10.1177/17470218221096148
qjep.sagepub.com
SAGE

Marcel R. Schreiner, Thorsten Meiser and Arndt Bröder

Abstract

Experienced events consist of several elements which need to be bound together in memory to represent the event in a coherent manner. Given such bindings, the retrieval of one event element should be related to the successful retrieval of another element of the same event, thus leading to a stochastic dependency of the retrieval of event elements. The way in which bindings are structured is not yet clearly established and only few moderators of the binding of event elements have been identified. We present results from three experiments aiming to distinguish between an integrated binding structure, in which event elements are bound into a unitary representation, and a hierarchical binding structure, in which event elements are preferentially bound to specific types of elements. Experiments 2 and 3 were additionally designed to identify animacy, an entity's property of being alive, as a potential moderator of the binding of event elements. We also offer a new approach for modelling dependencies of the retrieval of event elements which mitigates some limitations of previous approaches. Consistent with previous literature, we found dependencies of the retrieval of event elements if all of an event's constituent associations were shown. We found mixed evidence for integrated or hierarchical binding structures but found dependency of the retrieval of event elements to be sensitive to the presence of animacy in an event. The results suggest that binding structures may vary depending on moderators such as animacy or event structure awareness. Theoretical implications and directions for future research are discussed.

Keywords

Episodic memory; binding; animacy; statistical modelling

Received: 12 October 2021; revised: 8 February 2022; accepted: 23 March 2022

Episodic memory stores information about experienced events (Tulving, 1972, 1983) which consist of multiple elements, such as persons, objects, locations, actions, and sensations. Despite different event elements being represented in different neocortical regions (Alvarez & Squire, 1994; Horner et al., 2015), they need to be bound together to enable the retrieval of the event in a coherent manner. The hippocampus is considered to be the structure responsible for accomplishing this task (Backus et al., 2016; Cohen & Eichenbaum, 1993; Davachi et al., 2003; Eichenbaum et al., 2007; Squire & Zola-Morgan, 1991). Binding should be associated with an increased likelihood of retrieving subsequent event elements if a preceding element was successfully retrieved. This leads to a stochastic dependency of the retrieval of event elements (e.g., Arnold et al., 2019; Boywitt & Meiser, 2012a, 2012b; Bröder, 2009; Horner & Burgess, 2013, 2014; Horner et al., 2015; Meiser & Bröder, 2002; Ngo et al., 2019; Starns & Hicks, 2005, 2008) whereas it is not precluded that dependency is affected by retrieval-based processes in addition to binding processes occurring during encoding, such as suggested by Kumaran and McClelland (2012). However, there exist diverging views regarding the representational structure in which different event elements are bound together.

One purpose of the current research is to distinguish between an integrated binding structure, in which event elements are bound into a unitary representation, and a hierarchical binding structure, in which event elements are preferentially bound to particular elements. This relates to the fundamental principles driving information storage and retrieval in episodic memory. Some authors suggest that the hippocampus acts as a convergence zone, binding event elements into a single engram which can then be

Department of Psychology, School of Social Sciences, University of Mannheim, Mannheim, Germany

Corresponding author:
Marcel R. Schreiner, Department of Psychology, School of Social Sciences, University of Mannheim, L13, 15, 68161 Mannheim, Germany.
Email: m.schreiner@uni-mannheim.de

retrieved by partial activation of event elements via pattern completion (Damasio, 1989; Marr, 1971; Moll & Miikkulainen, 1997). This is consistent with Tulving's idea of event engrams as discrete bound event representations, containing information about different event elements (Tulving, 1983). A related view is integrative encoding, which suggests that the hippocampus integrates newly encountered associations into existing, overlapping, ones, ultimately leading to an integrated representation containing all event elements (Shohamy & Wagner, 2008; Zeithamova et al., 2012). We term such representations or engrams an integrated binding structure. Thus, in an integrated binding structure, elements of a given event and associations among these elements are represented in a single superordinate memory structure and can consequently be accessed in an all-or-none manner. However, results supporting integrative encoding may also be explained by pairwise, non-overlapping, representations of individual experiences (Kumaran & McClelland, 2012; McClelland et al., 1995). In addition, other views, such as ensemble encoding (Cai et al., 2016), relational memory theory (Cohen & Eichenbaum, 1993; Eichenbaum, 1999; see also Eichenbaum & Cohen, 1988, 2001), and the theory of event coding (TEC, Hommel et al., 2001), more strongly emphasise pairwise representations. Ensemble encoding posits that associations are stored as overlapping ensembles while remaining distinct rather than forming a unitary representation (Cai et al., 2016). Relational memory theory suggests that the hippocampus flexibly links event elements such that they can be recombined depending on task demands. In the TEC, codes of stimuli (feature codes) are activated upon perception and are then bound into so-called event files (Hommel, 1998, 2009). Event files do not consist of a unitary representation but rather of multiple local interconnections as a result of selective binding (Hommel, 1998, 2004). In addition, the degree to which feature codes contribute to the event file may vary (Hommel et al., 2001) and not all possible pairwise bindings are necessarily formed (Moeller et al., 2019). The TEC thus allows for an asymmetry of bindings in event files. This may explain findings of asymmetries in the retrieval of event elements, such that some types of elements serve as more effective retrieval cues or are retrieved more likely (e.g., Hayes et al., 2004; Nairne et al., 2017; Trinkler et al., 2006). Binding asymmetries are also possible in the recently proposed Span–Cospan model of episodic memory (Healy & Caudell, 2019). When events are presented as sequences of event segments, the model assumes that event elements form higher order representations of event segments which are represented by specific cells. The representations may consequently form further higher level representations up to a representation of the entire event while holistic access to individual event segments is maintained. Representations and connections can vary in strength. Thus, asymmetries are possible if the connection strength of cells responsible for representations at different levels varies such that certain combinations of event elements lead to stronger higher level representations. From these views, it follows that bindings may be hierarchically organised such that event elements are preferentially bound to one type of element. Thus, a hierarchical binding structure does not posit that event elements are represented in a unitary manner but rather that they are organised in a system of pairwise bindings in which some bindings may be systematically prioritised over others, allowing for asymmetries in binding strength.

The distinction between an integrated and a hierarchical binding structure is related to the discussion of the binding variability and the mutual cuing hypothesis in the source memory literature, which refers to memory for the conditions under which a memory has been acquired (Johnson et al., 1993). The binding variability hypothesis suggests that source features are primarily bound to the item rather than to each other (Starns & Hicks, 2005; see also the model of headed records, J. Morton et al., 1985), pointing to an item–feature hierarchy. The mutual cuing hypothesis suggests additional direct binding of features (Meiser & Bröder, 2002), which makes it more similar to an integrated binding structure. However, the mutual cuing hypothesis does not necessarily predict that item and features are bound into a unitary representation. There is an ongoing debate regarding the two accounts, with some results supporting mutual cuing (Boywitt & Meiser, 2012a, 2012b; Meiser & Bröder, 2002; see also Balaban et al., 2019) and others supporting binding variability (Hicks & Starns, 2016; Starns & Hicks, 2005, 2008; Vogt & Bröder, 2007). There is some evidence against an integrated binding structure in item-based representations as investigated in the source memory literature (Brady et al., 2013; Utochkin & Brady, 2020). Note, however, that item-based representations may differ from the more complex event-based representations that are the focus of the current research (Andermane et al., 2021; Brady et al., 2013; Joensen et al., 2020; Utochkin & Brady, 2020). Event-based representations consist of several elements, which can be considered to be item-based representations. Thus, item-based representations are nested within event-based representations (see Andermane et al., 2021). Item-based representations can also contain more specific information than event-based representations (Hunt & Einstein, 1981). Furthermore, event-based representations are potentially dynamic, include a spatiotemporal context, and allow for the construction of scenes, which is not the case for item-based representations (Andermane et al., 2021; Robin, 2018; Rubin & Umanath, 2015).

Direct behavioural evidence for integrated or hierarchical binding structures is scarce. Horner and Burgess (2013) found a dependency of the retrieval of event elements by having participants learn a series of events consisting of several elements (person, object, and location). For example,

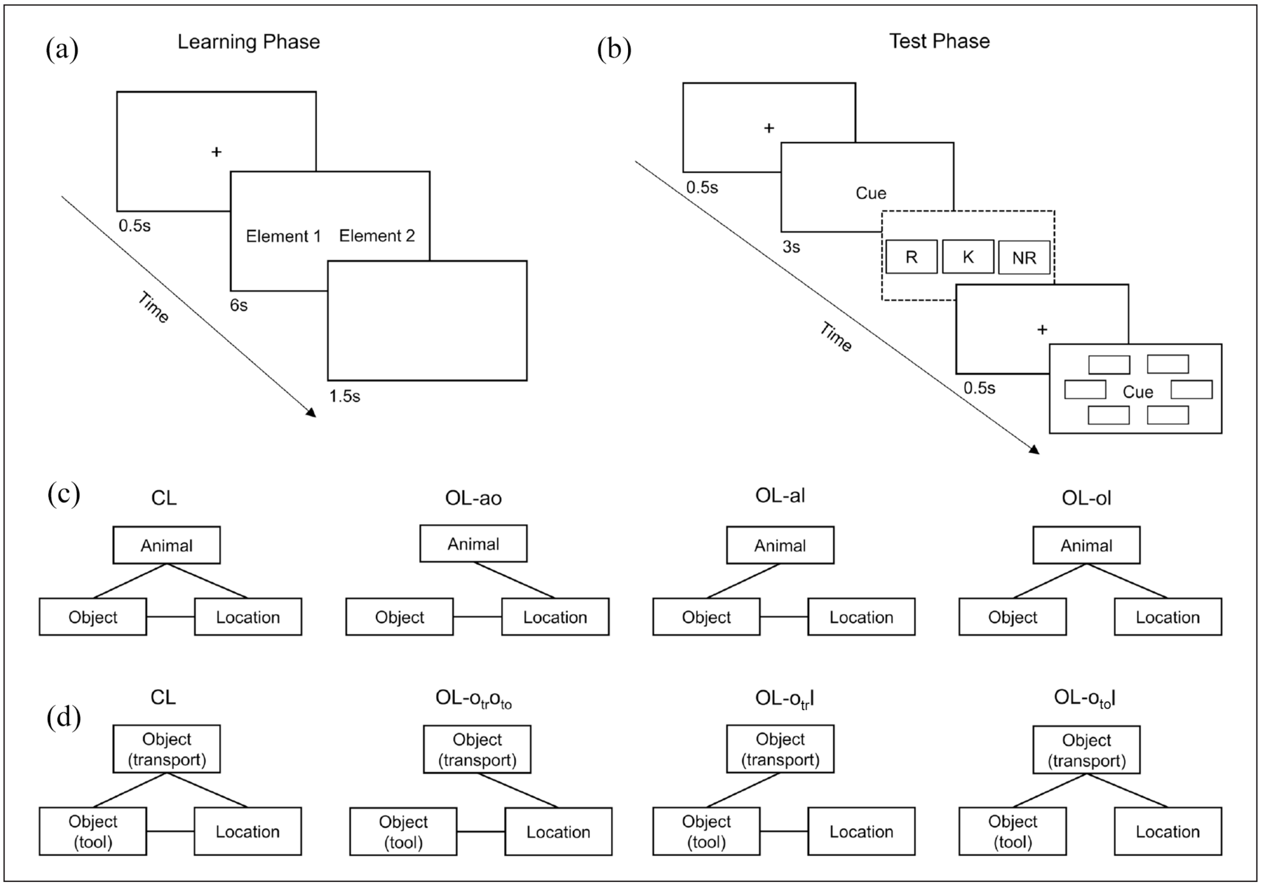


Figure 1. Experimental design and procedure. (a) Schematic depiction of a learning trial. (b) Schematic depiction of a test trial; recollection judgements (dashed rectangle) were only assessed in Experiment 1. (c) Associative structure of the experimental conditions in Experiment 1 and in the animacy condition of Experiments 2 and 3. (d) Associative structure of the non-animacy condition in Experiments 2 and 3.

Note. R = remember; K = know; NR = no recognition; CL = closed loop; OL-ao = open loop with association animal–object excluded; OL-al = open loop with association animal–location excluded; OL-ol = open loop with association object–location excluded; OL- $o_{tr}o_{to}$ = open loop with association means of transportation–tool excluded; OL- o_{tr}l = open loop with association means of transportation–location excluded; OL- o_{to}l = open loop with association tool–location excluded; transport = means of transportation.

participants may be presented *David Cameron–bicycle–swimming pool*. Horner and Burgess (2014) and Horner et al. (2015) built on this procedure and introduced the separated encoding paradigm in which each pairwise association is presented separately during encoding. For example, given the previous example event with the elements *David Cameron*, *bicycle*, and *swimming pool*, participants may be presented the pairs *David Cameron–bicycle*, *bicycle–swimming pool*, and *swimming pool–David Cameron* across different learning trials (see also Figure 1a). Note that in this paradigm, different learning trials referring to the same event are not presented in sequence but are interleaved with learning trials referring to other events. While this may deviate to some extent from how events are "naturally" experienced, it allows to manipulate the associative structure of an event presentation (see Horner et al., 2015). Dependency in a separated encoding condition was not reduced compared with simultaneous encoding (but see James et al., 2020, for boundary conditions). However, this was only the case when all events were presented in a closed-loop (CL) structure in which all pairwise associations are shown (i.e., all possible pairings of event elements), but not in an open-loop structure in which the presentation of one pairwise association is excluded such that, for example, *David Cameron–bicycle* is not presented (Horner & Burgess, 2014; Horner et al., 2015; see also Joensen et al., 2020). The authors concluded that binding depends on the coherence of the encoding episode. These results seem to be in favour of an integrated binding structure. However, the authors did not systematically vary the excluded association within the open-loop condition. Thus, the specific association being excluded in an event could vary within the open-loop condition. We argue that, for testing an integrated against a hierarchical binding structure, it is necessary to systematically vary the excluded association across different experimental conditions (see also Cabeza, 2006). If this is not done, associations that may be critical for binding are excluded for some events but not for others within the same condition. In addition, if associations are not excluded systematically, it may be the last presented association that yields coherence, as found by Horner et al. (2015). However, this may be different if associations are excluded systematically. Thus, in the current research, we focus on associations between event elements irrespective of presentation order. We used several open-loop conditions in each of which only one type of association (e.g., object–location) was excluded from presentation (see also Figure 1c and d) instead of a single open-loop condition in which the type of excluded association could vary. In addition, Horner and Burgess (2014) and Horner et al. (2015) used an approach for modelling stochastic dependencies of the retrieval of event elements introduced by Horner and Burgess (2013). This approach is based on contingency tables for the retrieval of event elements in different test pairs (i.e., pairs of test trials in a memory test), which are aggregated across events.

We propose a new approach for modelling the stochastic dependency of the retrieval of event elements based on item response theory (IRT; Lord, 1980; Lord & Novick, 1968) that takes individual item[1] responses as input. The approach exploits the assumption of local independence (LI) inherent in many IRT models. LI requires item responses to be independent given a general latent person trait such as memory performance (de Ayala, 2009; Lazarsfeld & Henry, 1968). If binding of event elements occurs, this would result in event-specific effects which influence item responses in addition to the general latent person trait. This would violate the LI assumption and manifest in nonzero residual correlations for item pairs belonging to the same event. The estimated item residual correlations are used for computing the dependency measure, which contrasts the item residual correlations within events with the item residual correlations between events. The approach provides several advantages over previous approaches such as the one by Horner and Burgess (2013) or Yule's *Q* (Yule, 1912; see also Hayman & Tulving, 1989; Horner & Burgess, 2014). It does not require the aggregation of responses into contingency tables and does not require the pre-specification of fixed test pairs, as is the case for previous approaches. In addition, our approach yields higher statistical power for detecting dependencies and differences in dependency between conditions than do previous approaches while providing good maintenance of Type I error rates (Schreiner & Meiser, 2022). Because previous approaches are based on aggregated contingency tables, they are prone to Simpson's paradox (Hintzman, 1972, 1980; Simpson, 1951), stating that collapsing 2×2 contingency tables into summary ones may lead to relationships of the two outcome variables in the summary tables diverge from the ones in the original tables. This is not the case for our approach because it is not contingency-based. In addition, our approach can account for varying item difficulties and allows to account for guessing. It can, in principle, also be applied to a greater variety of test formats such as free recall and is not limited to cued recall or cued recognition.

Based on the results of our first experiment, we additionally aimed to identify animacy as a potential moderator of the binding of event elements. Such moderators have largely been absent in the literature so far. To our knowledge, the results by James et al. (2020), which hint at the modality of stimulus presentation (written vs. pictorial) and the dimensionality of presentation modality (unimodal vs. multimodal) to be potential moderators of the binding of event elements in the context of the separated encoding paradigm (Horner & Burgess, 2014; Horner et al., 2015), are the only ones referring to this topic.

In the current research, we aim to determine whether event elements are bound into an integrated or a hierarchical structure and investigate animacy as a potential moderator in the binding of event elements. Building on the work by Horner and Burgess (2014) and Horner et al. (2015), we aim to overcome limitations of earlier studies by systematically varying the excluded associations and offering a novel approach for modelling the stochastic dependency of the retrieval of event elements which mitigates some limitations of previous approaches. To this end, we conducted three experiments. The results of Experiment 1 are in favour of a hierarchical binding structure in which event elements are preferentially bound to an animate element. Experiments 2 and 3 were designed to both replicate and extend the findings from Experiment 1 by additionally investigating whether animacy influences the binding of event elements. While the results in favour of a hierarchical binding structure did not replicate across experiments, the experiments yielded evidence that animacy influences the binding of event elements.

Experiment 1

In Experiment 1, we tested an integrated against a hierarchical binding account. We expected to replicate findings of a stochastic dependency of the retrieval of event elements (Hypothesis 1). In the source memory literature, a stochastic dependency of the retrieval of event elements has only been found for remember responses but not for know responses (Boywitt & Meiser, 2012a, 2012b; Meiser & Bröder, 2002; Meiser et al., 2008; Starns & Hicks, 2005). Remember and know responses are subjective ratings of memory quality, intended to tap into feelings of conscious recollection and experienced familiarity, respectively. While both recollection and familiarity enable recognition (Gardiner, 1988; Tulving, 1985), they may be different forms of memory with different functional characteristics (see Yonelinas, 2002, for a review). Similarly, we expected to only find a dependency of the retrieval of event elements in the case of recollection for event-based representations (Hypothesis 2). Previous findings suggest that dependency of the retrieval of event elements is eliminated if the encoding episode is not coherent (open-loop structure; Horner & Burgess, 2014; Horner et al., 2015). We suspected that effects may be masked because excluded associations in the open-loop condition were not systematically varied and due to limitations of the modelling approach used. While we expected that dependency is reduced in non-coherent encoding episodes, we did not expect that dependency is completely eliminated in such situations (Hypothesis 3). Finally, integrated and hierarchical binding structures make different predictions regarding dependency in non-coherent encoding episodes (open-loop structures), in which specific associations are excluded during study. An integrated binding structure suggests that the dependency does not vary as a function of the association being excluded. This is because an integrated binding structure consists of only a single unitary representation of the event that can be accessed in an all-or-none manner. Thus, the association that was not presented should readily be retrieved with the other associations from this unitary representation as if all associations were equally strong or retrieval should fail for all associations.[2] On the contrary, a hierarchical binding structure does not posit a unitary representation and it suggests an asymmetry in the binding strength of event elements, leading to some associations being more critical for dependency than others. Thus, excluding an association should affect more critical associations in some cases, so that stochastic dependency is diminished, and less critical associations in others, so that stochastic dependency is preserved or diminished to a smaller extent. Consequently, a hierarchical binding structure suggests that dependency varies as a function of the excluded association (Hypothesis 4). The experiment was preregistered at https://osf.io/ncpvq.

Method

Design. Each event consisted of the three constituent elements: animal, object, and location. There were four experimental within-subjects conditions (loop conditions). In the CL condition, all possible pairwise associations were presented (animal–object, animal–location, and object–location). In each of the three open-loop conditions, one pairwise association was consistently excluded from presentation (see also the paired-associate learning paradigm; e.g., Preston et al., 2004). Consequently, there was one condition in which animal–object was excluded (OL-ao), one in which animal–location was excluded (OL-al), and one in which object–location was excluded (OL-ol) (see Figure 1). Thus, events in the open-loop conditions consisted of two overlapping pairs with a common element. The design is an adaptation of the one used by Horner and Burgess (2014) and Horner et al. (2015) in the context of the separated encoding paradigm. We equated the open-loop conditions to the CL condition regarding the number of event elements instead of the number of associations. Previous research yielded similar results when equating the number of associations or event elements (Horner & Burgess, 2014; Joensen et al., 2020).

Material. Stimuli consisted of 180 German nouns of three different types—60 animals (all mammals; e.g., *dog*), common objects (e.g., *bucket*), and locations (e.g., *office*). An additional 12 nouns—four animals, common objects, and locations—were used as buffers to avoid primacy effects (primacy buffers). Stimuli were partly taken and adapted from the ones used by Joensen et al. (2020) and translated into German. We used animals instead of famous persons to prevent potential effects of prominence or ignorance of specific persons. From the stimuli, we randomly generated 60 animal–object–location triplets, making up an "event" for each participant. Events were then randomly assigned to the four experimental conditions, resulting in 15 events per condition. In addition, we randomly generated four primacy buffer events, one per condition, which were presented first.

Procedure. The experiment was conducted online and implemented using lab.js (Henninger et al., 2020). Data collection was managed by JATOS (Lange et al., 2015). The procedure (see Figure 1) was based on the separated encoding paradigm (Horner & Burgess, 2014; Horner et al., 2015). The experiment consisted of a learning phase, a filler phase, and a test phase. Participants were not made aware of the underlying event structure and were not informed that they would later be tested on the stimuli seen in the learning phase. In the learning phase, events were presented sequentially with two of the constituent elements (i.e., one association) shown per learning trial. There was a minimum of two other event trials between two same event trials. Words were presented to the left or right of the screen centre. The assignment of event element type (e.g., animal) to screen location was randomised. Participants were instructed to imagine the words as elements of a scene as vividly as possible and imagine them interacting in a meaningful manner. Each trial consisted of a 0.5-s fixation cross, followed by the presentation of the word pair for 6 s and a subsequent 1.5-s blank screen. The experimental conditions were randomly distributed across trials. Primacy buffer events were presented first to prevent primacy effects and were not included in the test phase. In the filler phase, participants had to solve randomly generated math problems for 3 min to avoid recency effects.

In the test phase, following a 0.5-s fixation cross, participants were first presented a cue word, which was an event element (e.g., an object) they had seen in the learning phase, in the screen centre for 3 s. Participants then had to give recollection judgements, indicating whether they *remembered* the cue word, merely *knew* that it had been presented in the learning phase or did *not recognise* it. This was done to distinguish between experiences of recollection and familiarity. The instructions for the remember–know distinction closely followed those used by Gardiner (1988), translated into German. Following another 0.5-s fixation cross, participants then conducted a cued recognition forced-choice task. The cue word was displayed in the screen centre, and response alternatives were displayed in a hexagonal array around it. Participants had to choose the target associated with the cue word from the response alternatives. All response alternatives were of the same type (e.g., location) and distractors were randomly drawn from other events. The screen location of the target was randomised. All associations were tested, but only in one direction to avoid testing effects. Thus, there were two possible configurations of cue–target pairs that could be tested for a given event: (a) cue animal and target object, cue object and target location, and cue location and target animal, and (b) cue animal and target location, cue location and target object, and cue object and target location. The direction tested was randomly determined per event, and thus each direction occurred, on average, equally often and randomly distributed across participants. This resulted in three test trials per event. Note that for the open-loop conditions, test trials included one inference trial per event in which the target and cue word were not presented jointly in the learning phase but belong to the same event. While they were not shown as being explicitly related, they could be flexibly related through their overlap with the common event element (for example, if participants learned the associations animal–object and animal–location, they may also imagine a relation between object–location and integrate it into a common memory representation). Thus, for inference trials, a correct response indicates a correct reconstruction of the association that was not shown in the learning phase. The test phase consisted of three blocks, with one association per event tested in each block. Trial order was randomised in each block. Thus, inference trials were intermixed with the other test trials.

Data analysis. All analyses were conducted in the R Programming Environment (R Core Team, 2020), and we used the R package *papaja* (Version 0.1.0.9997; Aust & Barth, 2020) for reporting. We used the conventional significance level of $\alpha = 5\%$ for all analyses.

Exploratory analysis of memory performance. To analyse memory performance, we fit a generalised linear mixed model with a logit link function (see Goldstein, 2011), using the test trial outcomes as a binary dependent variable. Note that the analysis refers to single trials and not aggregated values across trials (see Hoffman & Rovine, 2007). We included random person intercepts and fixed effects for condition, recollection judgement, association[3], and the interactions. To assess the influence of specific factors, we compared models with isolated effects with a baseline model. For the main effects, the baseline model was the null model that only contained a fixed and a random person intercept. For the two-way interactions, the baseline model was the model with all main effects, and for the three-way interaction, the baseline model was the model with all main effects and two-way interactions. For each effect, we then computed the Bayes factor in favour of an effect (BF_{10}) using Bayesian information criterion (BIC) approximation[4] (Raftery, 1995; Wagenmakers, 2007). Thus, a Bayes factor >1 is in favour of an effect. A Bayes factor >3 is considered moderate evidence, and a Bayes factor >10 is considered strong evidence for an effect (consequently, Bayes factors <0.33 and <0.1 are considered moderate and strong evidence for the absence of an effect, see Jeffreys, 1961). In addition, we computed the marginal pseudo-R^2 (Nakagawa et al., 2017), which describes the proportion of variance explained by the fixed effects, for each model and report the change in marginal R^2 (R^2_{change}) as an indicator of effect size. For the full model, we report both the marginal R^2 and the conditional R^2, which describes the proportion of variance explained by both fixed and random effects. To further investigate effects, we conducted post-hoc pairwise comparisons using the p value adjustment by Holm (1979) to account for multiple testing.

Models were fit using the R package *lme4* (Version 1.1-23; Bates et al., 2015). Pseudo-R^2 were computed using the package *MuMIn* (Version 1.43.17; Barton, 2020) using the delta method. Post-hoc pairwise comparisons were conducted using the package *emmeans* (Version 1.4.7; Lenth, 2020).

Analysis of dependency. To model the stochastic dependency of the retrieval of event elements, we employed an IRT (Lord, 1980; Lord & Novick, 1968) approach. Items (i.e., test trials in the cued recognition task, including inference trials) were ordered by condition, event, and cue type. We used a three-parameter logistic model (Birnbaum, 1968) because it allows to control for guessing. It models the probability of person i to give a correct response u to item j, given a latent person trait θ, the item difficulty β, an item-specific discrimination parameter γ, and an item-specific guessing parameter γ:

$$P\left(u_{\mathrm{ij}} = 1\right) = \gamma_{\mathrm{j}} + \left(1 - \gamma_{\mathrm{j}}\right) \frac{e^{\alpha_{\mathrm{j}}(\theta_i - \beta_{\mathrm{j}})}}{1 + e^{\alpha_{\mathrm{j}}(\theta_i - \beta_{\mathrm{j}})}} \tag{1}$$

As events were randomly generated, we fixed discrimination parameters to be equal across trials and set α_j to 1. We fixed guessing parameters to the stochastic guessing probability of $\frac{1}{6}$ given six response alternatives. This reduces the model to:

$$P\left(u_{\mathrm{ij}} = 1\right) = \frac{1}{6} + \frac{5}{6} \frac{e^{\theta_{\mathrm{i}} - \beta_{\mathrm{j}}}}{1 + e^{\theta_{\mathrm{i}} - \beta_{\mathrm{j}}}} \tag{2}$$

This model assumes LI of item responses, which means that the latent person trait, reflecting participants' memory performance, accounts for all inter-item relationships (de Ayala, 2009; Lazarsfeld & Henry, 1968). Consequently, the residual correlations between items should equal zero. This assumption is violated if there are other influences on item responses beyond the latent person trait. Given binding of event elements, there should be additional event-specific effects inducing a dependency of the retrieval of event elements within triplets over and above the dependency induced by the person effect θ. This would violate LI and manifest as nonzero residual correlations of related item pairs. We calculated item residual correlations using the Q_3 statistic (Yen, 1984). The statistic is calculated for item pairs (j, j') in four steps. First, person and item parameters are estimated from the model. Second, the probability of answering items j and j' correctly is determined for each person based on the estimated model parameters. Third, the residuals for both items are computed by subtracting the probability of a correct response from the observed response (i.e., 0 or 1) for each person. Finally, the Q_3 statistic for the item pair is calculated as the correlation of the residuals of both items across persons. Yen (1993) noted that the Q_3 statistic is negatively biased, with an expected value of $\frac{-1}{I-1}$ given LI, with I being the number of items. Thus, we applied a bias correction by subtracting this expected value from all Q_3. We defined the stochastic dependency of the retrieval of event elements (D) as:

$$D = \frac{1}{K}\sum_{k>k'} Q_3^{kk'} - \frac{1}{L}\sum_{l>l'} Q_3^{ll'} \tag{3}$$

where kk' are item pairs belonging to the same event, ll' are item pairs belonging to different events, K is the number of item pairs belonging to the same event, and L is the number of item pairs belonging to different events. Given stochastic dependency of the retrieval of event elements, within-event residual correlations should deviate from zero, whereas between-event residual correlations should not. Consequently, D should deviate from zero. Note that D is rather robust against model misspecification, because this affects both within- and between-event residual correlations. We calculated the dependency estimates for the whole data and for specific recollection judgements (remember, know, and no recognition responses).

Because the sampling distribution of Q_3, and consequently the sampling distribution of D, is unknown (Chen & Thissen, 1997), we obtained p values using parametric bootstrapping. To obtain estimates of event-specific effects to use in the parametric bootstrap, we fit a bifactor model (see Gibbons & Hedeker, 1992; Wainer & Wang, 2000; Supplementary Appendix A). The model extends the unidimensional IRT model in Equation 1 by adding additional latent traits for each event. Thus, items from one event load on one of these additional latent traits, and this trait is thus specific for a given event. These event-specific latent traits capture residual stochastic dependencies within the triplets forming an event. Stochastic dependencies are reflected by the traits' variances, with higher variances indicating higher stochastic dependencies within events. These variances can be used as indicators of event-specific effects in the parametric bootstrap. We restricted variances of event-specific traits to be equal within conditions, because events were randomly generated. We employed two different approaches. In the first approach, conditional independence depicts the null hypothesis, whereas residual dependencies between items of an event within a condition depict the alternative hypothesis. For this approach, we simulated 1,000 datasets from the unidimensional model in Equation 2. Item parameters were estimated from the data[5] and person parameters were drawn from a normal distribution with mean zero and variance estimated from the data. We then calculated D values for each dataset and recollection judgement, and computed two-tailed p values[6]. In the second approach, equal residual dependencies between conditions depict the null hypothesis, whereas differences in residual dependencies between conditions depict the alternative hypothesis. For this approach, we simulated 1,000 datasets per condition from the bifactor model in Equation A2 of Supplementary Appendix A. Item parameters were estimated from the data and person parameters were drawn from a multivariate normal distribution with a zero mean vector and variances estimated from the data. We set variances of all event-specific latent traits equal to the one estimated for the respective focal condition. For obtaining specific estimates for different recollection judgements, we assumed them to be randomly distributed across persons and items, with probability equalling their respective proportion in the data. We then calculated differences between D values and computed one-tailed p values for the differences[7]. Note that we did not test for differences in dependencies between pairs of conditions if there were negative dependencies in both conditions, because such a comparison is not relevant for the research questions. Further information on the modelling approach is given in Schreiner and Meiser (2022).

We used the R package *mirt* (Version 1.32.1; Chalmers, 2012) and adapted functions from the package *sirt* (Version 3.9-4; Robitzsch, 2020) for the dependency analysis. Simulations were conducted using the package *SimDesign* (Version 2.0.1; Chalmers, 2020). We also report the dependency results obtained using the approach by Horner and Burgess (2013) in Supplementary Appendix B. Results were largely congruent with the ones from the main dependency analysis.

Participants. Participants were recruited from the web (social media, mailing lists, forums, blogs, and the online research platform SurveyCircle) and could join a lottery for winning vouchers of a total value of 400€ and receive course credit (SurveyCircle, 2021). A power analysis using simulated data based on data from a pilot study ($n = 27$) for detecting the expected pattern of results with medium effects (differences in event-specific trait variances of 1

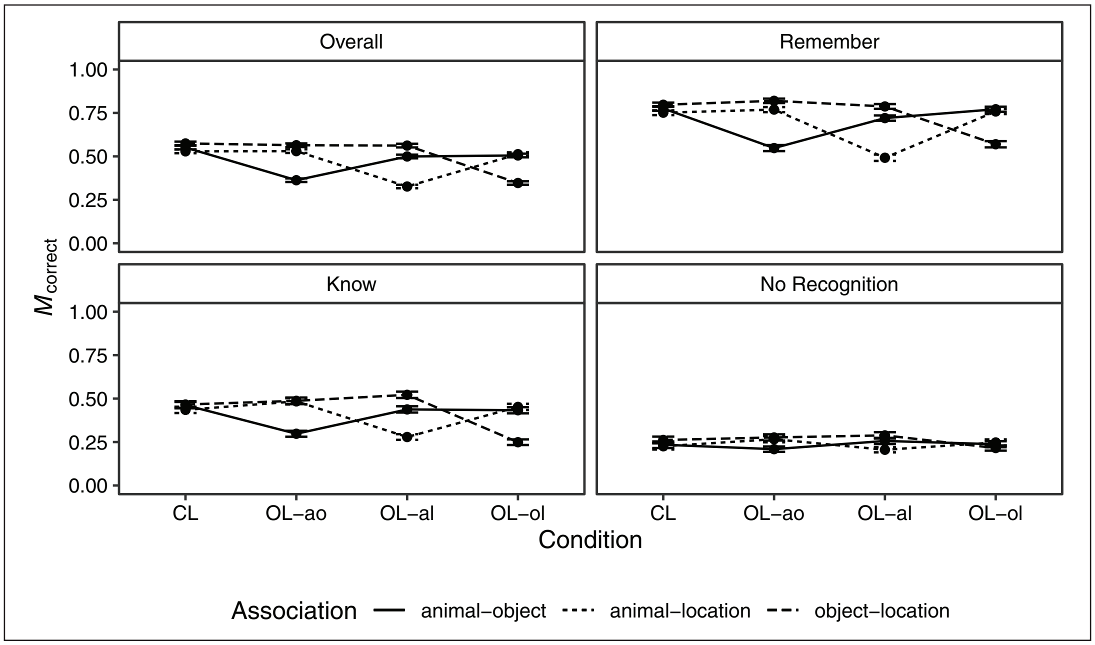


Figure 2. Mean proportion of correct responses by loop condition and association for the whole data (overall) and for subsets of data with specific recollection judgements in Experiment 1.

Note. CL = closed loop; OL-ao = open loop with association animal–object excluded; OL-al = open loop with association animal–location excluded; OL-ol = open loop with association object–location excluded. Error bars represent ±*SEM*.

according to the statistical procedure; cf. Glas et al., 2000; Wang et al., 2002) between conditions with 80% power (one-tailed testing) yielded a desired sample size of 180 participants. For further information about the power analysis, see Supplementary Appendix C. The experiment was completed by 181 participants. All participants provided online informed consent for their participation and publication of their data. One participant was excluded due to not speaking German fluently. Another participant was excluded due to low accuracy (less than 10%) in the filler task. Another four participants were excluded because they indicated their data should not be used (e.g., due to missing some learning trials). Two additional participants were excluded because they indicated having recently participated in a similar study. Finally, 24 participants were excluded because they interrupted the experiment[8]. This yielded a final sample of 149 participants[9] (72% female, 1% non-binary, 1% not wanting to disclose their gender; 75% students) with a mean age of 27.0 years ($SD = 8.5$). Data, materials, and analysis scripts for the experiment are provided via the Open Science Framework (OSF; https://osf.io/dt35k/).

Results

Memory performance. Overall, the proportion of correct responses was $M = 0.49$ ($SD = 0.50$). The proportion of correct responses by condition, association, and recollection judgement is shown in Figure 2. Further indices are shown in Table D1 in the Supplementary Appendix. There was strong evidence for main effects of condition ($\text{BF}_{10} > 1{,}000$, $R^2_{\text{change}} = .007$), recollection judgement ($\text{BF}_{10} > 1{,}000$, $R^2_{\text{change}} = .07$), and association ($\text{BF}_{10} > 1{,}000$, $R^2_{\text{change}} = .001$). Post-hoc pairwise comparisons revealed significantly higher performance for remember responses than for know (log-odds ratio [log OR][10] = 1.02, $z = 27.14$, $p < .001$) and no recognition responses (log OR = 1.46, $z = 33.98$, $p < .001$), and significantly higher performance for know than for no recognition responses (log OR = 0.44, $z = 10.79$, $p < .001$). There was also strong evidence for a two-way interaction of condition and association ($\text{BF}_{10} > 1{,}000$, $R^2_{\text{change}} = .03$) which qualified the respective main effects. Post-hoc pairwise comparisons (see Table 1) revealed that memory performance was lowest in conditions in which the respective association was not presented in the learning phase (i.e., inference associations) but did not significantly differ otherwise, except for lower performance for the association animal–object than for object–location in condition OL-al. In condition CL, performance was lower for association animal–location than for object–location but did not significantly differ otherwise. There was strong evidence against two-way interactions of condition and recollection judgement ($\text{BF}_{10} < 0.001$, $R^2_{\text{change}} = .001$) and of recollection judgement and association ($\text{BF}_{10} < 0.001$, $R^2_{\text{change}} < .001$). Finally, there was strong evidence against a three-way interaction ($\text{BF}_{10} = 0.02$, $R^2_{\text{change}} = .002$). The marginal R^2 of the full model was .11 and the conditional R^2 was .33.

Table 1. Results of the post-hoc pairwise comparisons for the interaction of condition and association regarding memory performance in Experiment 1.

Contrast	Condition	Log OR	*z*	*p*
Animal-object–animal-location	CL	0.13	1.71	.26
Animal-object–object-location	CL	–0.12	–1.63	.26
Animal-location–object-location	CL	–0.25	–3.34	.004
Animal-object–animal-location	OL-ao	–0.90	–12.18	<.001
Animal-object–object-location	OL-ao	–1.06	–14.29	<.001
Animal-location–object-location	OL-ao	–0.17	–2.26	.09
Animal-object–animal-location	OL-al	0.89	12.10	<.001
Animal-object–object-location	OL-al	–0.34	–4.72	<.001
Animal-location–object-location	OL-al	–1.23	–16.63	<.001
Animal-object–animal-location	OL-ol	–0.03	–0.47	.64
Animal-object–object-location	OL-ol	0.84	11.26	<.001
Animal-location–object-location	OL-ol	0.88	11.79	<.001

Note. Log OR = log-odds ratio; CL = closed loop; OL-ao = open loop with association animal–object excluded; OL-al = open loop with association animal–location excluded; OL-ol = open loop with association object–location excluded.

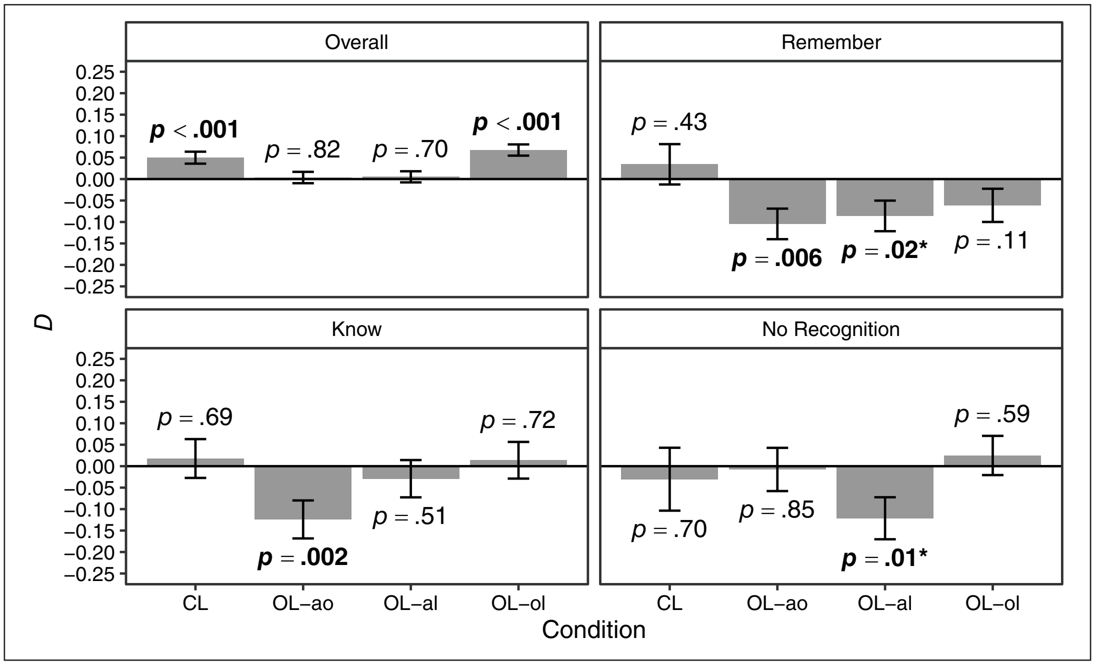


Figure 3. Dependency of the retrieval of event elements by loop condition in Experiment 1 for the whole data (overall) and for subsets of data with specific recollection judgements.

Note. CL = closed loop; OL-ao = open loop with association animal–object excluded; OL-al = open loop with association animal–location excluded; OL-ol = open loop with association object–location excluded. Error bars represent ±*SE*. The *p* values set in boldface indicate statistical significance at the $p < .05$ level; *p* values marked with an asterisk (*) did no longer indicate statistical significance at the $p < .05$ level after adjusting for multiple comparisons using the *p* value adjustment by Holm (1979); *p* values were obtained using parametric bootstrapping.

Dependency. Dependencies of the retrieval of event elements are shown in Figure 3. Overall, there was a significant positive dependency in conditions CL and OL-ol but not in conditions OL-ao and OL-al. The dependency in condition CL was significantly larger than the dependency in condition OL-ao ($D_{\text{diff}} = 0.05$, $p = .04$), although this difference was no longer significant after adjusting for multiple comparisons ($P_{\text{adj}} = .14$) using the *p* value adjustment by Holm (1979). The dependency in condition CL did not significantly differ from the dependencies in conditions OL-al ($D_{\text{diff}} = 0.04$, $p = .06$) and OL-ol ($D_{\text{diff}} = –0.02$, $p = .13$). The dependencies in conditions OL-ao and OL-al did not significantly differ ($D_{\text{diff}} = 0.00$, $p = .43$) but were significantly smaller than the dependency in condition OL-ol ($D_{\text{diff}} = –0.06$, $p = .001$ and $D_{\text{diff}} = –0.06$, $p < .001$, respectively). Regarding specific recollection judgements, there were significant negative dependencies for the subset of the data that received remember responses in conditions OL-ao and OL-al, although the latter was no longer significant after adjusting for multiple comparisons ($p_{\text{adj}} = .07$). The dependency in condition CL was significantly larger than the dependencies in the open-loop conditions ($D_{\text{diff}} = 0.14$, $p = .005$; $D_{\text{diff}} = 0.12$, $p = .02$; and $D_{\text{diff}} = 0.10$, $p = .04$, respectively). Regarding know responses, there was a significant negative dependency in condition OL-ao. The dependency in condition CL was significantly larger than the dependency in condition OL-ao ($D_{\text{diff}} = 0.14$, $p = .02$) but did not significantly differ from the dependencies in conditions OL-al ($D_{\text{diff}} = 0.05$, $p = .29$) and OL-ol ($D_{\text{diff}} = 0.00$, $p = .49$). The dependency in condition OL-ao was significantly smaller than the dependency in condition OL-ol ($D_{\text{diff}} = -0.14$, $p = .03$), although this difference was no longer significant after adjusting for multiple comparisons ($p_{\text{adj}} = .11$). The dependency in condition OL-al did not significantly differ from the dependency in condition OL-ol ($D_{\text{diff}} = -0.04$, $p = .27$). Regarding no recognition responses, there was a significant negative dependency in condition OL-al, although this dependency was no longer significant after adjusting for multiple comparisons ($p_{\text{adj}} = .06$). The dependency in condition OL-al was significantly smaller than the dependency in condition OL-ol ($D_{\text{diff}} = -0.15$, $p = .03$), although this difference was no longer significant after adjusting for multiple comparisons ($p_{\text{adj}} = .09$). The dependencies in conditions CL and OL-ao did not significantly differ from the dependency in condition OL-ol ($D_{\text{diff}} = -0.06$, $p = .25$ and $D_{\text{diff}} = -0.03$, $p = .34$, respectively). In summary, there were significant positive dependencies in the CL condition and in the open-loop condition in which the association object–location was excluded. These dependencies were significantly larger than the ones in the other open-loop conditions, in which dependencies did not significantly differ from zero. Dependencies for specific recollection judgements did either not significantly differ from zero or were significantly negative.

Discussion

In Experiment 1, we replicated the finding that the dependency of the retrieval of event elements is maintained if the encoding of an event occurs in several temporally divided episodes (Horner & Burgess, 2014; Horner et al., 2015; Joensen et al., 2020). Thus, Hypothesis 1, which stated that there is a stochastic dependency of the retrieval of event elements, was supported.

Hypothesis 2, which stated that dependency is only found in the case of recollection, was not supported. Dependency was not only and not consistently found for remember responses. Results regarding specific recollection judgements were inconsistent, and if dependencies reached significance, they were unexpectedly negative. It is also noteworthy that the dependency pattern for the whole data differed considerably from the dependency patterns for specific recollection judgements. This may be because the overall pattern also includes dependencies between item responses associated with different recollection judgements. These are excluded when only considering item responses associated with specific recollection judgements. For example, relationships between event elements may be remembered better for remember than for know responses. This may also extend to item pairs where one item received a remember response and the other received a know response (i.e., despite one item receiving a know response, all relationships are remembered well). However, responses to such item pairs are only considered when using the whole data but not when estimating the dependency for remember or know responses in isolation. The inconsistent findings regarding specific recollection judgements may suggest that the remember–know paradigm in its current implementation is not appropriate for use together with the separated encoding paradigm and the more complex representations studied. The remember–know paradigm (Gardiner, 1988) targets only specific elements. This is appropriate for simpler representations, such as an object with two features. As we closely adapted the paradigm for the current experiment, recollection judgements refer to specific cue words. However, the separated encoding paradigm and the modelling approach operate on the level of associations and whole events. It may be this discrepancy in targeting levels that drives the inconsistent findings regarding specific recollection judgements. Another potential limitation may be participants struggling to understand the remember–know instructions (e.g., see Geraci et al., 2009; Migo et al., 2012), which may limit the validity of the subjective remember–know responses. In addition, differences in information contributing to the dependency estimates for different recollection judgements (i.e., varying number of item responses considered in the computation of the respective estimates) and differences in memory performance associated with different recollection judgements may have limited the equatability of estimates for different recollection judgements, which may have contributed to the unexpected findings. However, we considered these differences in the parametric bootstrap, and thus, the significance patterns of the empirical results should be comparable for different recollection judgements.

Hypothesis 3 stated that dependency is reduced but not eliminated in non-coherent encoding episodes. Dependency was effectively eliminated in conditions OL-ao and OL-al, although dependency in condition CL was not significantly larger than the dependency in condition OL-al and not significantly larger than the dependency in condition OL-ao after adjusting for multiple comparisons. However, this may be due to a power problem. Also note that adjusting *p* values is associated with a loss of statistical power. Nevertheless, the tests against independence clearly support the interpretation that dependency was effectively eliminated in these conditions. In condition OL-ol, however, dependency was maintained and did not significantly differ from the dependency in condition CL. Thus, Hypothesis 3 was not supported. However, this pattern of results supports Hypothesis 4, which stated that dependency varies as a function of the excluded association in non-coherent encoding episodes. Excluding the association object–location in the learning phase did not affect dependency, whereas excluding associations involving the animal did. This was the case even though the pairwise associations did generally not differ regarding memory performance given that they were shown in the learning phase. The pattern of results suggests a hierarchical binding structure in which elements are preferentially bound to the animal. In addition, the results suggest that the encoding episode does not necessarily have to be coherent for dependencies to occur. In Experiment 2, we aimed to replicate these findings and determine whether the observed pattern of results can be attributed to animacy influencing the binding of event elements.

Experiment 2

Human memory functioning may be a product of selective pressure on our ancestors (Nairne et al., 2007, 2008). In this context, animacy may be an especially important survival-related factor influencing human cognition (Nairne et al., 2013, 2017). For example, words with an animate referent are retrieved more likely than words with an inanimate referent, a phenomenon termed the animacy effect (e.g., Li et al., 2016; Nairne et al., 2013; VanArsdall et al., 2015). Such an animacy effect has been found for several types of tasks such as free recall (Bonin et al., 2015; Li et al., 2016; Madan, 2021; Nairne et al., 2013; Popp & Serra, 2016), cued recall (DeYoung & Serra, 2021; Laurino & Kaczer, 2019; VanArsdall et al., 2015; but note Kazanas et al., 2020; Popp & Serra, 2016, who found reduced performance for animate referents in cued recall tasks), free recognition (Bonin et al., 2014; see also VanArsdall et al., 2013), and judgements of learning (DeYoung & Serra, 2021; Li et al., 2016). Animate entities are defined as being living things which are capable of independent movement and can change direction without warning (Bonin et al., 2015). The animal event elements in Experiment 1 meet this definition. Given that the results of Experiment 1 suggest that elements are preferentially bound to the animal and the importance of animacy in human cognition (Nairne et al., 2013, 2017), it may be that animacy affects not only the retrieval but also the binding of event elements. For example, animacy may qualify the referent word to be an initiator of action, thus qualifying it to be the grammatical subject in sentences describing events, whereas inanimate objects or locations are grammatical objects.

In Experiment 2, we aimed to investigate whether animacy was responsible for the effect found in Experiment 1. To this end, we constructed events that either include an animate element, as was the case in Experiment 1, or do not include an animate element. If animacy is responsible for the effect in Experiment 1, the dependency of the retrieval of event elements should vary as a function of the excluded association in non-coherent encoding episodes if events include an animate element (Hypothesis 5a). Specifically, for these events, the pattern of results of Experiment 1 should be replicated. However, the dependency of the retrieval of event elements should not vary as a function of the excluded association in non-coherent encoding episodes if events do not include an animate element (Hypothesis 5b). We decided not to further investigate dependency for different recollection judgements, but instead focus on the main research questions of how the binding of event elements in episodic memory is structured and whether animacy influences binding. The experiment was preregistered at https://osf.io/m2fjv.

Method

Design. Half of the events included an animate entity and the other half did not, leading to a 2 (animacy condition: animacy vs. non-animacy) × 4 (loop condition: CL and three open loops) within-subjects design. For the animacy condition, loop conditions were identical to those of Experiment 1. In the open-loop non-animacy conditions, the association means of transportation–tool (OL-$o_{tr}o_{to}$), means of transportation–location (OL-$o_{tr}l$), or means of tool–location (OL-$o_{to}l$) was excluded from presentation (see Figure 1).

Material. Stimuli consisted of 192 German nouns, partly taken from Experiment 1, of four different types—32 animals (all mammals), 48 objects representing means of transportation (e.g., *bicycle*), 48 objects representing tools (e.g., *hammer*), and 64 locations. An additional 24 nouns—four animals, six means of transportation, six tools, and eight locations—were used as primacy buffers. From the stimuli, we randomly created 64 triplets, making up an "event" for each participant. Half of the events consisted of an animal, an object (balanced as to whether being a means of transportation or a tool), and a location (animacy condition). The other half consisted of two objects (one means of transportation and one tool) and a location (non-animacy condition). Events were then randomly assigned to the eight experimental conditions, resulting in eight events per condition.[11] In addition, we randomly generated eight primacy buffer events, one per condition, which were presented first.

Procedure. The procedure was identical to the one of Experiment 1 with the following exceptions: For each participant, stimuli were kept separate for the animacy and non-animacy conditions to keep the number of possible distractors in the test phase equal between different types of elements. To achieve this, one-third of the means of transportation and the tools stimuli were initially randomly assigned to the animacy condition, while the remaining ones were used for the non-animacy condition. In addition, we did not collect recollection judgements in this experiment. Thus, a test trial only consisted of a 0.5-s fixation cross, followed by a 3-s cue presentation, followed by another 0.5-s fixation cross, followed by the cued recognition task.

Data analysis. Data analysis was identical to the one conducted in Experiment 1 except that we did not consider recollection judgements in this experiment. For the exploratory analysis of memory performance, we included loop condition, animacy condition, association, and the interactions as fixed effects in the generalised linear mixed model. We coerced the associations animal–object and means of transportation–tool, animal–location and means of transportation–location, and object–location and tool–location into a common factor level, respectively. We also coerced loop conditions OL-ao and OL-$o_{tr}o_{to}$, OL-al and OL-o_{tr}l, and OL-ol and OL-o_{to}l into a common factor level, respectively.[12] For the dependency analysis, p values were again obtained using parametric bootstrapping.

Participants. Participants were recruited from the web and could receive course credit or a monetary compensation of 3€ and join a lottery for winning vouchers of a total value of 100€. A power analysis using simulated data based on Experiment 1 for detecting the expected pattern of results with small to medium effects (differences in event-specific trait variances of 0.75; cf. Glas et al., 2000; Wang et al., 2002) between conditions with 80% power (one-tailed testing) yielded a desired sample size of 210 participants. Given the observed exclusion rate in Experiment 1, we decided to increase the desired sample size by 20% and thus collected data of 252 participants. All participants provided online informed consent for their participation and publication of their data. Two participants were excluded due to not speaking German fluently. Another two participants were excluded due to low accuracy (less than 10%) in the filler task. Another three participants were excluded because they indicated their data should not be used (e.g., due to distractions). Two additional participants were excluded because they indicated having recently participated in a similar study. Finally, 30 participants were excluded because they interrupted the experiment. This yielded a final sample of 213 participants (73% female, 0.5% non-binary, 1% not wanting to disclose their gender; 80% students) with a mean age of 27.3 years ($SD=9.5$). Data, materials, and analysis scripts for the experiment are provided via the OSF (https://osf.io/dt35k/).

Results

Memory performance. Overall, the proportion of correct responses was $M=0.40$ ($SD=0.49$) in the animacy condition and $M=0.38$ ($SD=0.49$) in the non-animacy condition. The proportion of correct responses by loop condition, animacy condition, and association is shown in Figure 4a. Further indices are shown in Table D2 in the Supplementary Appendix. There was strong evidence for a main effect of loop condition ($\mathrm{BF}_{10} > 1{,}000$, $R^2_{\mathrm{change}} = .005$) but weak evidence against a main effect of animacy condition ($\mathrm{BF}_{10} = 0.59$, $R^2_{\mathrm{change}} < .001$) and strong evidence against a main effect of association ($\mathrm{BF}_{10} = 0.02$, $R^2_{\mathrm{change}} < .001$). There was strong evidence for a two-way interaction of loop condition and association ($\mathrm{BF}_{10} > 1{,}000$, $R^2_{\mathrm{change}} = .04$) which qualified the main effect of loop condition. Post-hoc pairwise comparisons (see Table 2) revealed that memory performance was lowest in conditions in which the respective association was not presented in the learning phase (i.e., inference associations) but did not significantly differ otherwise. In condition CL, performance was lower for association animal–object/transport–tool than for object–location/tool–location but did not significantly differ otherwise. There was strong evidence against two-way interactions of loop condition and animacy condition ($\mathrm{BF}_{10} < 0.001$, $R^2_{\mathrm{change}} < .001$) and of animacy condition and association ($\mathrm{BF}_{10} = 0.003$, $R^2_{\mathrm{change}} < .001$). Finally, there was strong evidence against a three-way interaction ($\mathrm{BF}_{10} < 0.001$, $R^2_{\mathrm{change}} < .001$). The marginal R^2 of the full model was .04 and the conditional R^2 was .28.

Dependency. Dependencies of the retrieval of event elements are shown in Figure 5a. There were no significant dependencies in all conditions except for a negative dependency in loop condition OL-$o_{tr}o_{to}$ in the non-animacy condition, which was no longer significant after adjusting for multiple comparisons ($p_{\mathrm{adj}} = .06$) using the p value adjustment by Holm (1979). The dependency in loop condition CL in the non-animacy condition was significantly larger than the dependency in loop condition OL-$o_{tr}o_{to}$ ($D_{\mathrm{diff}} = 0.06$, $p = .02$). All other relevant differences were non-significant ($p \geqslant .10$).

Discussion

In Experiment 2, we did not find evidence for substantial dependencies of the retrieval of event elements. We could neither replicate the positive dependency in condition CL nor the positive dependency in condition OL-ol from Experiment 1. Thus, the results cannot properly distinguish between an integrated and a hierarchical binding structure. As the pattern of results was similar for the animacy and non-animacy condition, there was also no evidence for a special role of animacy. Contrary to Experiment 1, in which events consisted of an animal, an object, and a

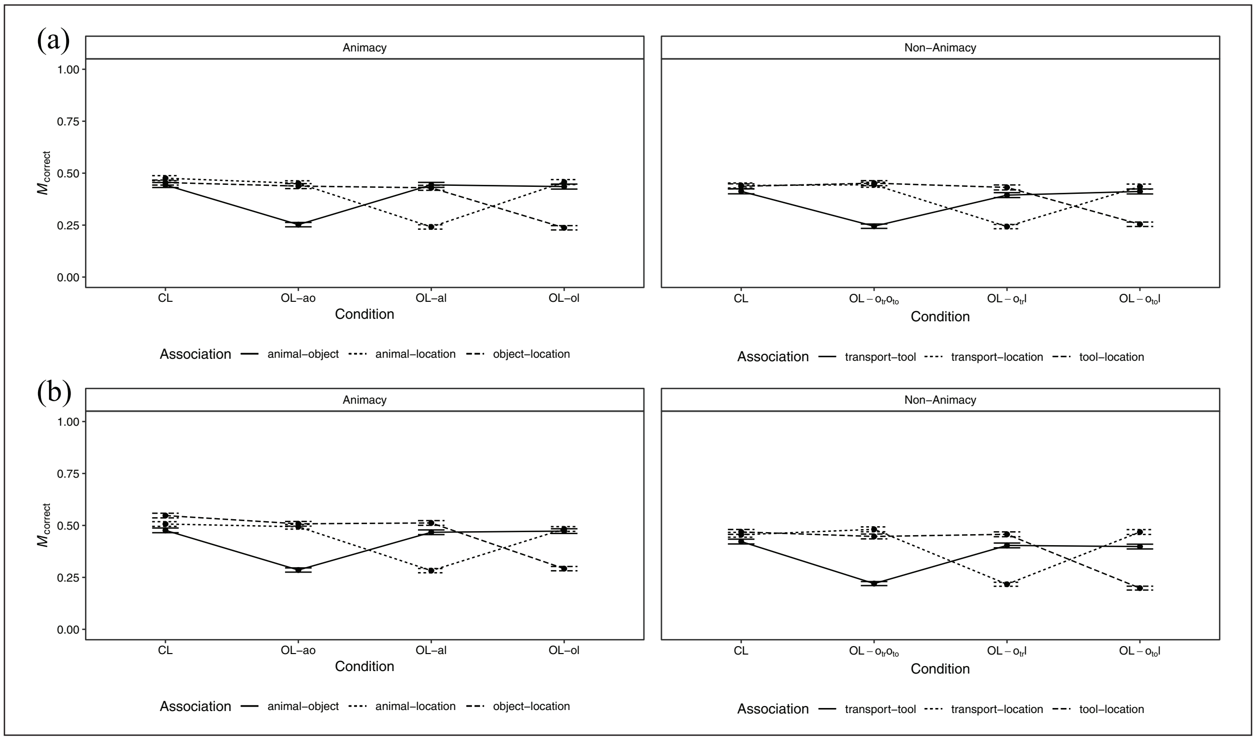


Figure 4. Mean proportion of correct responses by animacy condition, loop condition, and association in (a) Experiment 2 and (b) Experiment 3.

Note. CL = closed loop; OL-ao = open loop with association animal–object excluded; OL-al = open loop with association animal–location excluded; OL-ol = open loop with association object–location excluded; OL- $o_{tr}o_{to}$ = open loop with association means of transportation–tool excluded; OL- $o_{tr}l$ = open loop with association means of transportation–location excluded; OL- $o_{to}l$ = open loop with association tool–location excluded; transport = means of transportation. Error bars represent ±*SEM*.

location, the event structure in Experiment 2 was not always the same. Due to the full within-subjects design and the inclusion of a non-animacy condition, events could either consist of an animal, an object, and a location, or of two objects and a location. As these different event structures were presented in randomly alternating sequence, this may have more strongly concealed the underlying event structure. Thus, participants may not have been as aware of the event structures as in Experiment 1, preventing them from forming abstract representations of event structures, which may have caused them to use different encoding strategies (cf. N. W. Morton et al., 2020; see also Kumaran & Ludwig, 2013). For example, N. W. Morton et al. (2020) suggested that the formation of abstract event structures facilitates binding and particularly supports inference. The results do not preclude a hierarchical binding structure with animal as the critical element, but the varying event structures due to the full within-subjects design may have prevented the formation of coherent memory structures. This could be an additional moderator which requires further examination. In addition, the number of events per condition was reduced from 15 in Experiment 1 to 8 in Experiment 2. Thus, the condition-specific results are based on less information than in Experiment 1. To make the experimental design more similar to Experiment 1 and to rule out potential influences of different degrees of event structure awareness caused by varying event structures, in Experiment 3 we varied animacy as a between-subjects instead of a within-subjects factor and increased the number of events per condition back to 15.

Experiment 3

In Experiment 3, we again aimed to investigate whether animacy was responsible for the effect found in Experiment 1, while avoiding potential confounds which may have been present in Experiment 2. Thus, we varied animacy as a between-subjects factor and used the same number of events per condition as in Experiment 1. The experiment was preregistered at https://osf.io/vprxd.

Method

Design. The experimental design was identical to the one of Experiment 2 with the exception that animacy was manipulated as a between-subjects instead of a within-subjects factor. This resulted in a 2 (animacy condition: animacy vs. non-animacy) × 4 (loop condition: CL and three open loops) mixed design.

Table 2. Results of the post-hoc pairwise comparisons for the interaction of loop condition and association regarding memory performance in Experiment 2.

Contrast	Loop condition	Log OR	*z*	*p*
Animal-object/transport-tool–animal-location/transport-location	CL	–0.16	–2.89	.02
Animal-object/transport-tool–object-location/tool-location	CL	–0.09	–1.71	.35
Animal-location/transport-location–object-location/tool-location	CL	0.07	1.19	.71
Animal-object/transport-tool–animal-location/transport-location	OL-ao/$o_{tr}o_{to}$	–1.17	–19.48	<.001
Animal-object/transport-tool–object-location/tool-location	OL-ao/$o_{tr}o_{to}$	–1.15	–19.22	<.001
Animal-location/transport-location–object-location/tool-location	OL-ao/$o_{tr}o_{to}$	0.02	0.28	.78
Animal-object/transport-tool–animal-location/transport-location	OL-al/ $o_{tr}l$	1.06	17.52	<.001
Animal-object/transport-tool–object-location/tool-location	OL-al/ $o_{tr}l$	–0.06	–1.11	.71
Animal-location/transport-location–object-location/tool-location	OL-al/ $o_{tr}l$	–1.12	–18.58	<.001
Animal-object/transport-tool–animal-location/transport-location	OL-ol/ $o_{to}l$	–0.12	–2.15	.16
Animal-object/transport-tool–object-location/tool-location	OL-ol/ $o_{to}l$	1.06	17.61	<.001
Animal-location/transport-location–object-location/tool-location	OL-ol/ $o_{to}l$	1.18	19.64	<.001

Note. Log OR = log-odds ratio; CL = closed loop; OL-ao = open loop with association animal–object excluded; OL-al = open loop with association animal–location excluded; OL-ol = open loop with association object–location excluded; OL-$o_{tr}o_{to}$ = open loop with association means of transportation–tool excluded; OL-$o_{tr}l$ = open loop with association means of transportation–location excluded; OL-$o_{to}l$ = open loop with association tool–location excluded; transport = means of transportation. Associations and loop conditions separated by a slash (/) were treated as one factor level, respectively.

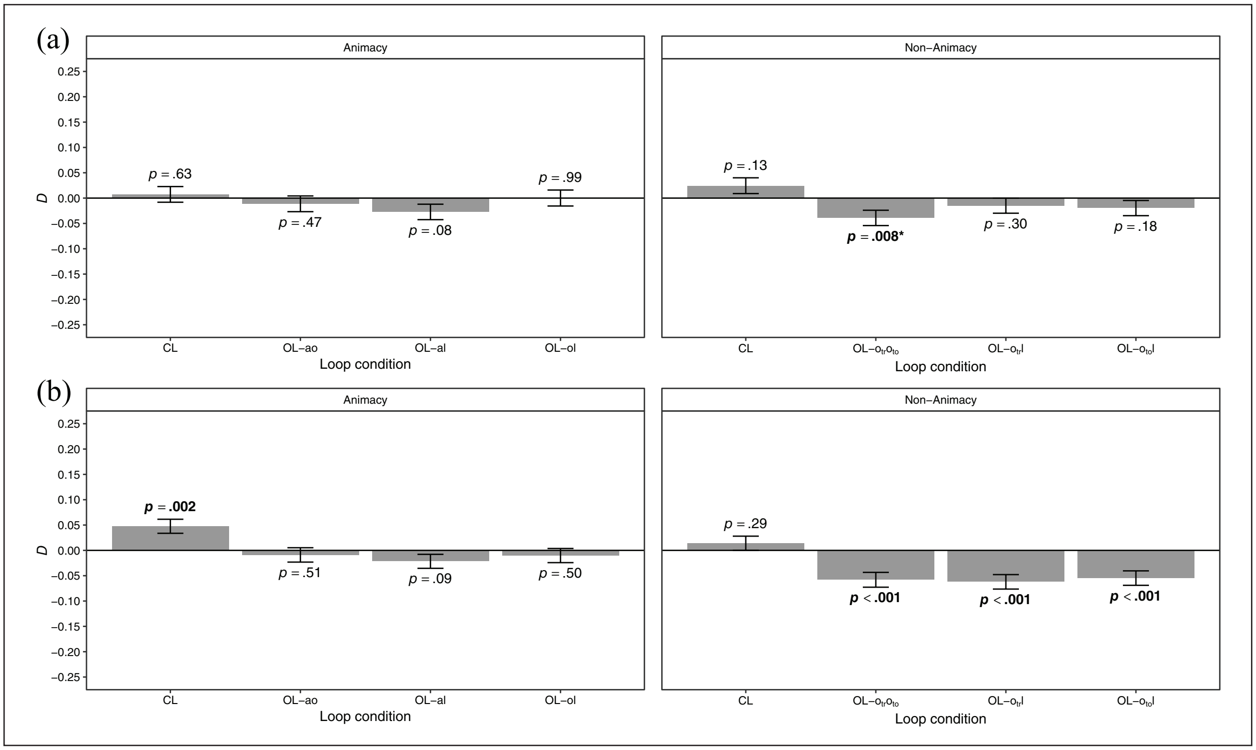


Figure 5. Dependency of the retrieval of event elements in the animacy and non-animacy conditions of (a) Experiment 2 and (b) Experiment 3 by loop condition.

Note. CL = closed loop; OL-ao = open loop with association animal–object excluded; OL-al = open loop with association animal–location excluded; OL-ol = open loop with association object–location excluded; OL- $o_{tr}o_{to}$ = open loop with association means of transportation–tool excluded; OL- $o_{tr}l$ = open loop with association means of transportation–location excluded; OL- $o_{to}l$ = open loop with association tool–location excluded. Error bars represent ±*SE*. The *p* values set in boldface indicate statistical significance at the $p < .05$ level; *p* values marked with an asterisk (*) did no longer indicate statistical significance at the $p < .05$ level after adjusting for multiple comparisons using the *p* value adjustment by Holm (1979); *p* values were obtained using parametric bootstrapping.

Material. Stimuli consisted of 240 German nouns, partly taken from Experiments 1 and 2, of four different types—60 animals (all mammals), 60 objects representing means of transportation, 60 objects representing tools, and 60 locations. An additional 16 nouns—four animals, means of transportation, tools, and locations—were used as primacy buffers. From the stimuli, we randomly created 60 triplets, making up an "event" for each participant. In the animacy condition, events consisted of an animal, an object (balanced as to whether being a means of transportation or a tool), and a location. In the non-animacy condition, events consisted of two objects (one means of transportation and one tool) and a location. Events were then randomly assigned to the four within-subjects conditions, resulting in 15 events per loop condition. In addition, we randomly generated four primacy buffer events, one per loop condition, which were presented first.

Procedure. The procedure was identical to the one of Experiment 2. In the animacy condition, for each participant, 30 means of transportation and 30 tools were randomly drawn from the respective lists to serve as object elements.

Data analysis. Data analysis was identical to the one conducted in Experiment 2 except that we used animacy condition as a between-subjects factor in the exploratory analysis of memory performance and fit separate models to the data of each animacy condition for the dependency analysis. For the dependency analysis, p values were again obtained using parametric bootstrapping.

Participants. Participants were recruited from the web and could join a lottery for winning vouchers of a total value of 450€ and earn course credit. A power analysis using simulated data based on Experiment 1 for detecting the expected pattern of results with medium effects (differences in event-specific trait variances of 1; cf. Glas et al., 2000; Wang et al., 2002) between conditions with 80% power (one-tailed testing) yielded a desired sample size of 260 participants (130 per between-subjects condition). Given the observed exclusion rate in Experiment 2, we decided to increase the desired sample size by 15%, and thus collected data of 299 participants (152 in the animacy condition and 147 in the non-animacy condition). All participants provided online informed consent for their participation and publication of their data. Five participants were excluded due to not speaking German fluently. Another four participants were excluded due to low accuracy (less than 10%) in the filler task. Another 10 participants were excluded because they indicated their data should not be used (e.g., due to technical problems or distractions). Four additional participants were excluded because they indicated having recently participated in a similar study. Finally, 23 participants were excluded because they interrupted the experiment. This yielded a final sample of 253 participants (131 in the animacy condition and 122 in the non-animacy condition; 75% female, 1.6% non-binary, 1% not wanting to disclose their gender; 81% students) with a mean age of 27.2 years ($SD = 9.1$). Data, materials, and analysis scripts for the experiment are provided via the OSF (https://osf.io/dt35k/).

Results

Memory performance. Overall, the proportion of correct responses was $M = 0.44$ ($SD = 0.50$) in the animacy condition and $M = 0.39$ ($SD = 0.49$) in the non-animacy condition. The proportion of correct responses by loop condition, animacy condition, and association is shown in Figure 4b. Further indices are shown in Table D3 in the Supplementary Appendix. There was strong evidence for a main effect of loop condition ($\mathrm{BF}_{10} > 1{,}000$, $R^2_{\text{change}} = .007$) and of association ($\mathrm{BF}_{10} > 1{,}000$, $R^2_{\text{change}} = .001$), but strong evidence against a main effect of animacy condition ($\mathrm{BF}_{10} = 0.04$, $R^2_{\text{change}} = .004$). There was strong evidence for a two-way interaction of loop condition and association ($\mathrm{BF}_{10} > 1{,}000$, $R^2_{\text{change}} = .05$) which qualified the respective main effects. Post-hoc pairwise comparisons (see Table 3) revealed that memory performance was lowest in conditions in which the respective association was not presented in the learning phase (i.e., inference associations). In addition, performance was lower for association animal–object/transport–tool than for object–location/tool–location in condition OL-al and lower for association animal–object/transport–tool than for animal–location/transport–location in condition OL-ol. In condition CL, performance was highest for association object–location/tool–location and lowest for association animal–object/transport–tool. Other comparisons were not significant. There was strong evidence against two-way interactions of loop condition and animacy condition ($\mathrm{BF}_{10} < 0.001$, $R^2_{\text{change}} < .001$) and of animacy condition and association ($\mathrm{BF}_{10} = 0.02$, $R^2_{\text{change}} < .001$). Finally, there was strong evidence against a three-way interaction ($\mathrm{BF}_{10} < 0.001$, $R^2_{\text{change}} = .001$). The marginal R^2 of the full model was .06 and the conditional R^2 was .27.

Dependency. Dependencies of the retrieval of event elements are shown in Figure 5b. In the animacy condition, there was a significant positive dependency in condition CL but no significant dependencies in the open-loop conditions. The dependency in condition CL was significantly larger than the dependencies in the open-loop conditions ($D_{\text{diff}} = 0.06$, $p = .007$; $D_{\text{diff}} = 0.07$, $p = .003$; and $D_{\text{diff}} = 0.06$, $p = .01$, respectively). In the non-animacy condition, there was no significant dependency in condition CL but significant negative dependencies in the open-loop conditions. The dependency in condition CL was significantly larger than the dependencies in the open-loop conditions ($D_{\text{diff}} = 0.07$, $p = .001$; $D_{\text{diff}} = 0.08$, $p < .001$; and $D_{\text{diff}} = 0.07$, $p = .002$, respectively).

Discussion

In Experiment 3, we could replicate the positive dependency in condition CL in the animacy condition, thus supporting Hypothesis 1, which stated that there is a stochastic dependency of the retrieval of event elements. Dependencies

Table 3. Results of the post-hoc pairwise comparisons for the interaction of loop condition and association regarding memory performance in Experiment 3.

Contrast	Loop condition	Log OR	*z*	*p*
Animal-object/transport-tool–animal-location/transport-location	CL	–0.16	–3.12	.005
Animal-object/transport-tool–object-location/tool-location	CL	–0.30	–5.82	<.001
Animal-location/transport-location–object-location/tool-location	CL	–0.14	–2.71	.01
Animal-object/transport-tool–animal-location/transport-location	OL-ao/ $o_{tr}\, o_{to}$	–1.30	–23.41	<.001
Animal-object/transport-tool–object-location/tool-location	OL-ao/ $o_{tr}\, o_{to}$	–1.26	–22.54	<.001
Animal-location/transport-location–object-location/tool-location	OL-ao/ $o_{tr}\, o_{to}$	0.05	0.92	.36
Animal-object/transport-tool–animal-location/transport-location	OL-al/ $o_{tr}l$	1.06	18.98	<.001
Animal-object/transport-tool–object-location/tool-location	OL-al/ $o_{tr}l$	–0.25	–4.83	<.001
Animal-location/transport-location–object-location/tool-location	OL-al/ $o_{tr}l$	–1.31	–23.51	<.001
Animal-object/transport-tool–animal-location/transport-location	OL-ol/ $o_{to}l$	–0.20	–3.88	<.001
Animal-object/transport-tool–object-location/tool-location	OL-ol/ $o_{to}l$	1.10	19.50	<.001
Animal-location/transport-location–object-location/tool-location	OL-ol/ $o_{to}l$	1.30	23.13	<.001

Note. Log OR = log-odds ratio; CL = closed loop; OL-ao = open loop with association animal–object excluded; OL-al = open loop with association animal–location excluded; OL-ol = open loop with association object–location excluded; OL-$o_{tr}o_{to}$ = open loop with association means of transportation–tool excluded; OL- $o_{tr}l$ = open loop with association means of transportation–location excluded; OL- $o_{to}l$ = open loop with association tool–location excluded; transport = means of transportation. Associations and loop conditions separated by a slash (/) were treated as one factor level, respectively.

were close to zero in the open-loop conditions in the animacy condition and negative in the open-loop conditions in the non-animacy condition. Thus, Hypothesis 3, which stated that dependency is reduced but not eliminated in non-coherent encoding episodes, was not supported. The negative dependencies in the non-animacy condition indicate that successful retrieval of one event element is associated with a decreased likelihood of retrieving another event element of the same event. One explanation for this may be that learning trials were encoded as distinct overlapping events. Zotow et al. (2020) found negative dependencies in such a case and suggested that they may be due to pattern separation processes in the hippocampus driving individual event representations apart. Another explanation may be retrieval-induced forgetting (Anderson et al., 1994). The selective retrieval of an event element (e.g., tool when cued by location) may inhibit the non-tested element (e.g., means of transportation), which is then retrieved less likely in the subsequent test trial in which it is the target (cf. Horner & Burgess, 2013). This may have particularly occurred in the non-animacy condition, because it contained two element types, means of transportation and tools, for which object could be considered a superordinate category. Thus, means of transportation and tools may be considered to be more similar semantic categories than, for example, animal and object, which may have facilitated retrieval-induced forgetting (cf. Hicks & Starns, 2004).

We could not replicate the positive dependency in condition OL-ol which was observed in Experiment 1. Thus, Hypothesis 4, which stated that dependency varies as a function of the excluded association in non-coherent encoding episodes, was not supported in Experiment 3, and the pattern of results is in favour of an integrated binding structure. Hypotheses 5a and 5b stated that dependency varies as a function of the excluded association in non-coherent encoding episodes if events include an animate element, but does not vary if evens do not include an animate element. While dependencies in the open-loop conditions in the non-animacy condition were very similar, thus supporting Hypothesis 5b, they were negative. In addition, dependencies did not vary across the open-loop conditions in the animacy condition. Thus, Hypothesis 5a was not supported. However, the results still suggest that animacy influences the binding of event elements. Rather than characterising the element to which other event elements are preferentially bound, as implied by Hypotheses 5a and 5b, the results suggest that animacy facilitates the binding of event elements if the encoding episode is coherent. In the absence of animacy, this integration seems to be less successful, as indicated by the non-significant dependency in condition CL in the non-animacy condition. In addition, if animacy is not present in an event and the event is encoded as temporally divided episodes, the different learning trials may be encoded as distinct events.

Effect of presentation order regarding animacy

We only observed positive stochastic dependencies of the retrieval of event elements for events that include an animate element (i.e., an animal in the current experiments) across experiments. This may be because animacy provides a potential agent in an event, which may facilitate the formation of coherent memory representations. Consequently, dependencies may be larger for events for which an association involving an animal (i.e., animal–object or animal–location) was presented first compared with events for which an association not involving an

Table 4. Dependency for events for which an association involving an animate element was presented first or not first per experiment and condition.

Experiment	Condition	Animate element first	*D*	*p*
1	CL	Yes	0.06	.002
1	CL	No	0.03	.26
1	OL-ao	Yes	0.00	.83
1	OL-ao	No	0.00	.97
1	OL-al	Yes	0.01	.53
1	OL-al	No	0.01	.75
2	CL	Yes	–0.01	.56
2	CL	No	0.03	.28
2	OL-ao	Yes	–0.02	.51
2	OL-ao	No	–0.01	.68
2	OL-al	Yes	–0.05	.02
2	OL-al	No	0.00	.97
3	CL	Yes	0.07	<.001
3	CL	No	–0.01	.90
3	OL-ao	Yes	–0.01	.47
3	OL-ao	No	0.00	.81
3	OL-al	Yes	–0.04	.06
3	OL-al	No	0.00	.81

Note. CL = closed loop; OL-ao = open loop with association animal–object excluded; OL-al = open loop with association animal–location excluded. Only animacy conditions excluding condition OL-ol were considered.

animal (i.e., object–location) was presented first. To examine whether this interpretation may be valid, we conducted an exploratory post hoc analysis of presentation order regarding animacy.

We computed dependencies separately for events for which an association involving an animal was presented first and for events for which it was not by declaring respective responses as missing values and then fitting separate models for the two cases. For this analysis, we only considered the animacy conditions, excluding condition OL-ol because in this condition only associations involving an animal were presented. For the bootstrap, we used estimates from the main models but declared some event responses as missing values based on the proportion of events for which an association involving an animal was presented first or not first in each experiment and considered condition.

The results are shown in Table 4. Of the conditions that yielded significant positive dependencies in the main dependency analyses (condition CL in Experiments 1 and 3)[13], we only found significant dependencies for events for which an association involving an animal was presented first, but not for events for which an association not involving an animal was presented first. This is in favour of the interpretation that the presence of an animate element in an event facilitates the formation of coherent memory representations by providing a potential agent.

General discussion

The purpose of this research was to determine whether event elements in episodic memory are bound in an integrated or a hierarchical manner and, based on the results of the first experiment, investigate whether the presence of animacy in an event influences the binding of its constituent elements, while introducing a new approach for modelling dependencies of the retrieval of event elements in episodic memory. The results of this research cannot clearly distinguish between an integrated and a hierarchical binding structure. However, they provide evidence that animacy influences the binding of event elements. In addition, they hint at a role of awareness regarding the structure of event elements in the binding of event elements.

In two out of three experiments, we found a positive stochastic dependency of the retrieval of event elements in coherent encoding episodes (closed-loop structures) if one of the event elements was animate. This is consistent with the previous literature (Horner & Burgess, 2014; Horner et al., 2015; James et al., 2020; Ngo et al., 2019) and supports Hypothesis 1. It indicates that event elements are bound together even if an event is experienced as several temporally divided encoding episodes. We did not find this effect in Experiment 2, in which events could take different structures for the same participant. In addition, encoding episodes referring to events with different structures were presented in randomly alternating sequence. Thus, the underlying event structure, while being implicit in all experiments, was likely harder for participants to determine in Experiment 2. This reduced awareness regarding the structure of event elements may have prevented participants from forming abstract representations of event structures (cf. N. W. Morton et al., 2020; see also Kumaran

& Ludwig, 2013) and may have caused them to use different encoding strategies in Experiment 2 compared with Experiments 1 and 3. The results thus hint at a moderating influence of event structure awareness on the binding of event elements, which may be influenced by perceived task demands. This is consistent with relational memory theory (Cohen & Eichenbaum, 1993; Eichenbaum, 1999), which suggests that task demands affect the binding of event elements. Interestingly, Horner and Burgess (2014) and Horner et al. (2015) also varied event structures and still found a significant dependency of the retrieval of event elements. In their experiments, each element type appeared equally often. This was not the case in our Experiment 2, in which there were fewer animals than means of transportation and tools (the two object categories used) and fewer means of transportation and tools than locations. In addition, their experiments encompassed fewer events than ours (36 events compared with 64 events in Experiment 2), which may have reduced participants' memory load compared with our experiments. These factors may have contributed to an increased awareness regarding event structures in the experiments by Horner and Burgess (2014) and Horner et al. (2015) compared with Experiment 2.

We also investigated how the binding of event elements differs regarding different recollection judgements. Whereas past research has only observed stochastic dependencies of the retrieval of event elements for remember responses but not for know responses (Boywitt & Meiser, 2012a, 2012b; Meiser & Bröder, 2002; Meiser et al., 2008; Starns & Hicks, 2005), the present study did not find a consistent pattern across different recollection judgements, and dependencies were mostly unexpectedly negative. There is thus no support for Hypothesis 2. However, the remember–know paradigm (Gardiner, 1988) was usually used in the context of item-based representations and targets only specific cue elements. In the context of more complex event-based representations (cf. Andermane et al., 2021; Joensen et al., 2020), which were the focus of the current research, this leads to a discrepancy in targeting levels between the remember–know paradigm and the experimental paradigm and modelling approach, because the latter operate on the level of associations and whole events. This discrepancy may explain the inconsistent findings regarding recollection judgements. The results suggest that the remember–know paradigm may not be readily transferable to more complex representations, at least not in the form of our adaptation of the paradigm.

Regarding non-coherent encoding episodes (open-loop structures), dependencies were either close to zero or not reduced compared with coherent encoding episodes, at least if events contained an animate element. The results do not support Hypothesis 3 but are partly consistent with previous research, which found dependencies only in coherent but not in non-coherent encoding episodes (Horner & Burgess, 2014; Horner et al., 2015; Joensen et al., 2020). Indeed, non-coherent encoding episodes seem to generally disrupt the formation of coherent memory representations, as indicated by the absence of dependencies, or facilitate the formation of pairwise bindings, as opposed to higher level binding structures, which then exhibit mutual inhibition or suppression, as may be indicated by the negative dependencies in conditions in which events did not contain an animate element. Potential mechanisms behind negative dependencies may include pattern separation processes in the hippocampus, which drive individual representations apart (cf. Zotow et al., 2020), or retrieval-induced forgetting (Anderson et al., 1994).

The finding that dependency varied across the non-coherent encoding conditions in Experiment 1, with higher dependency if the association object–location than the associations animal–object or animal–location were excluded from the learning phase, supports Hypothesis 4 and suggests a hierarchical binding structure in which event elements are preferentially bound to the animal. This contradicts previous accounts and interpretations of the binding of event elements as being integrative, such as accounts advocating that event elements are bound into a single coherent event representation or engram (Damasio, 1989; Horner & Burgess, 2014; Horner et al., 2015; Joensen et al., 2020; Marr, 1971; Moll & Miikkulainen, 1997; Tulving, 1983) and the integrative encoding hypothesis (Shohamy & Wagner, 2008; Zeithamova et al., 2012). Rather, the finding is consistent with accounts considering asymmetrical binding such as the ensemble encoding account (Cai et al., 2016), relational memory theory (Cohen & Eichenbaum, 1993; Eichenbaum, 1999), the TEC (Hommel et al., 2001), and the Span–Cospan model of episodic memory (Healy & Caudell, 2019). However, in Experiment 3, the dependency in the non-coherent encoding condition with association object–location being excluded could not be replicated. Dependencies in the non-coherent encoding episodes were all close to zero. This is in favour of an integrated binding structure and thus consistent with integrative binding accounts (Damasio, 1989; Horner & Burgess, 2014; Horner et al., 2015; Joensen et al., 2020; Marr, 1971; Moll & Miikkulainen, 1997; Shohamy & Wagner, 2008; Tulving, 1983; Zeithamova et al., 2012). The results of Experiment 2 are not diagnostic for distinguishing between an integrated and a hierarchical binding structure because even the established finding of a dependency in the coherent encoding condition was not replicated. Taken together, evidence for Hypothesis 4 is ambiguous, and thus the results do not clearly distinguish between an integrated and a hierarchical binding structure. It may well be the case that both integrated and hierarchical binding structures are possible, with the binding structure formed determined by several moderators. James et al. (2020) already identified the modality of stimulus presentation and the dimensionality of presentation modality as potential moderators of the binding of event elements in the context of the separated encoding paradigm (Horner & Burgess, 2014; Horner et al., 2015).

Another moderator may be animacy (e.g., see Bonin et al., 2015; Nairne et al., 2013, 2017). In the current research, positive stochastic dependencies have only been observed for events that include an animate element. However, in Experiment 3, in which events with an animate element and events without any animate element were directly contrasted, dependency did not vary across the non-coherent encoding conditions and was even negative for events without an animate element. These results do not support Hypotheses 5a and 5b but still suggest an influence of animacy. Rather than characterising the prominent event element in a hierarchical binding structure, animacy seems to facilitate the binding of event elements per se, at least in the case of coherent encoding episodes. Although dependencies could also result from processes occurring during retrieval rather than encoding (e.g., Kumaran & McClelland, 2012), we would argue that *differences* in the stochastic dependencies of the retrieval of event elements between animacy conditions imply that there are also differences in the internal representations of the events between the conditions. We prefer to interpret these representation differences in terms of "binding" because this provides a coherent interpretation, but other theoretical ideas may also be viable. The negative dependencies (i.e., successful retrieval of an event element being associated with reduced probability to retrieve another event element) found for events without an animate element may be due to retrieval-induced forgetting (Anderson et al., 1994). Another explanation may be that the temporally divided encoding episodes are represented as distinct overlapping events, thus consisting of pairwise bindings. Zotow et al. (2020) found negative dependencies for partially overlapping events and attributed these to pattern separation processes in the hippocampus which drive representations apart, decreasing their similarity. One could argue that negative dependencies may also occur due to between-event binding of event elements, for example, due to the prevalence of systematic conjunction errors (e.g., Reinitz et al., 1992). This was not the case in the experiments because mean between-event residual correlations were very close to zero in all conditions and experiments. Animacy may provide structure to an event by providing a potential agent. This may enable encoding strategies such as representing the event as a sentence, with the agent as the grammatical subject. In the absence of a prominent agent, events may not be as clearly structured and such encoding strategies not as easily applicable. Consequently, people may resort to pairwise bindings (see Cai et al., 2016; Cohen & Eichenbaum, 1993; Eichenbaum, 1999). In favour of this interpretation, we found significant positive dependencies when only considering events for which an association involving an animate element was presented first but not when only considering events for which an association not involving an animate element was presented first for conditions in which there was a significant positive dependency.

Importantly, our findings cannot be attributed to differences in memory performance between conditions. Memory performance did, with few exceptions, not vary across conditions. Unsurprisingly, memory performance was lower for to-be-inferred associations in the open-loop conditions, resulting in an overall higher performance in the CL conditions in which all associations were shown in the learning phase. We did not find a difference in memory performance between events that include an animate element and events that do not. On the level of associations, there were generally also no differences between associations involving an animate element and associations not involving an animate element. Memory performance for associations not involving an animate element even tended to be higher in some conditions. We did thus not find an animacy effect in terms of memory performance. While the effect has been shown using a variety of test formats such as free recall (Bonin et al., 2015; Li et al., 2016; Madan, 2021; Nairne et al., 2013; Popp & Serra, 2016; VanArsdall et al., 2015), cued recall (DeYoung & Serra, 2021; Laurino & Kaczer, 2019), free recognition (Bonin et al., 2014; see also VanArsdall et al., 2013), and judgements of learning (DeYoung & Serra, 2021; Laurino & Kaczer, 2019), results using cued recall have been mixed (DeYoung & Serra, 2021; Kazanas et al., 2020; Laurino & Kaczer, 2019; Popp & Serra, 2016) and the effect has not yet been examined in the context of cued recognition tests which we used in the current research. In addition, Bonin et al. (2015) found that an imagery instruction improves performance for inanimate words but not for animate words. As we instructed participants to imagine the presented words as elements of a scene and to imagine them interacting in a meaningful manner, this instruction may have prevented the emergence of an animacy effect regarding memory performance by boosting memory performance for the inanimate elements. Considering the diluting effect of mental imagery on animacy effects, the potency of animacy in influencing the binding of event elements may actually be underestimated in Experiments 2 and 3.

Taken together, our findings suggest that binding structures may change depending on event characteristics and perceived task demands. While they do not clearly distinguish between an integrated and a hierarchical binding structure, they suggest animacy to influence the binding of event elements and hint at an influence of event structure awareness.

Limitations

There are at least three potential limitations concerning the current research. First, due to the COVID-19 pandemic and the resulting limitations regarding lab-based data collection, all experiments were conducted online and took about 45 to 65 min to complete. Web-based studies naturally do not have the degree of experimental control that can be achieved in lab-based studies. However, several studies have shown comparable data quality for web- and lab-based studies (Armitage & Eerola, 2020; Bartneck et al., 2015; Dandurand et al., 2008; de Leeuw & Motz, 2016; Hilbig, 2016). A decrease in attention is also not necessarily found in web-based studies (Clifford & Jerit, 2014; Hauser & Schwarz, 2016) and the precision of stimulus timing of lab.js (Henninger et al., 2020), which was used for the implementation of our experiments, was found to be good (Anwyl-Irvine et al., 2021; Bridges et al., 2020). In addition, James et al. (2020) used the separated encoding paradigm in a web-based format before and found highly replicable effects. We too found the effect of a positive dependency when the encoding episode is coherent and events include an animate element in two out of three experiments, which is in favour of the robustness of the effect in web-based settings and sufficient data quality in our experiments.

Second, the separated encoding paradigm (Horner & Burgess, 2014; Horner et al., 2015) deviates to some extent from how events are "naturally" experienced, because temporal dependencies between event segments are reduced due to the interleaved presentation of learning trials referring to different events. However, the paradigm allows to manipulate the associative structure of event presentations, which is necessary when trying to distinguish between different binding structures, which was one of the goals of the current research. In addition, it allows to explore, for example, presentation order effects, such as whether dependency is higher for events for which an association involving an animate element was presented first than for events for which an association not involving an animate element was presented first.

Third, while we believe the newly proposed approach for modelling dependencies of the retrieval of event elements to be a substantial improvement over existing approaches, it has some limitations. First, it is somewhat limited in terms of the type of comparisons that can be conducted. Because the sampling distribution of the dependency index is unknown, it requires bootstrapping to draw statistical inferences. Thus, when comparing dependency indices of different conditions, only pairwise comparisons are currently possible. Second, floor or ceiling effects of memory performance may lead to an unreliable estimation of dependency indices, a problem that is also inherent to other measures. The results of the memory performance analysis, however, indicate that this was not an issue in our experiments. Third, if there are items that have no variance, the estimation of item parameters for these items is not possible. The risk of this to occur increases with smaller samples and more missing values. However, this was also not an issue in the current research. Fourth, while the modelling approach is rather robust against model misspecifications, model misspecifications may nevertheless lead to small shifts in dependency estimates and obtained p values. The same may be true for different sorting of items due to variability in item parameter estimation. When using parametric bootstrapping to obtain p values, these are to some degree also affected by Monte Carlo error. The Monte Carlo error can be reduced by increasing the number of bootstrap samples. We recommend to use at least 1,000 bootstrap samples (cf. Davison & Hinkley, 1997).

Directions for future research

In terms of future research, it is necessary to conduct additional studies to obtain evidence distinguishing between an integrated and a hierarchical binding structure. We think that the separated encoding paradigm (Horner & Burgess, 2014; Horner et al., 2015) with systematic variations of the excluded associations as done in the current research is a useful paradigm to this end. It may not necessarily be the case that binding always occurs in the same way. On the contrary, our results suggest that binding may be influenced by several moderators. We deem it very important to identify and clarify such moderators in future research, a topic that is yet underrepresented in the literature. Identifying these moderators will help to exert more experimental control and to rule out additional explanations for observed or unobserved effects. As our results hinted at a role of awareness of event structures in the binding of event elements, future research could examine effects of varying event structures or task demands systematically. In addition, the role of animacy in the binding of event elements should be examined more closely. For example, if animacy exerts its role by making available an agent in the event, agency instead of animacy may be causal for the effects. Consequently, similar effects should be found when manipulating the agency of specific event elements. It may also prove fruitful to manipulate presentation order (i.e., whether an association involving an animate element is presented first or an association not involving an animate element is presented first) systematically, because our post hoc analysis on this matter suggested an effect of presentation order. Furthermore, because the results regarding specific recollection judgements were quite inconsistent, future research could try different adaptations of the remember–know paradigm to evaluate its suitability for more complex representations such as those that are the focus of the current research. Finally, the newly proposed approach for modelling dependencies of the retrieval of event elements warrants further systematic examination to identify other potential strengths and weaknesses and areas for improvement.

Conclusion

In three experiments, we investigated whether the binding of event elements in episodic memory occurs in an integrated manner, in which event elements are bound into a unitary representation, or in a hierarchical manner, in which event elements are preferentially bound to particular elements. The experiments yielded inconsistent results which cannot clearly distinguish between an integrated and a hierarchical binding structure, which necessitates further research. However, the experiments yielded evidence that animacy influences the binding of event elements, a moderator that has not been previously considered. In addition, we identified event structure awareness, which may be affected by variability in event structure, as a potential additional moderator. Thus, the binding of event elements may vary based on several moderators such as animacy and perceived task demands. Finally, we provide a new approach for modelling dependencies of the retrieval of event elements in episodic memory which mitigates some limitations of previous approaches.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—GRK 2277 "Statistical Modeling in Psychology."

ORCID iD

Marcel R. Schreiner https://orcid.org/0000-0001-7719-6281

Data accessibility statement

The data, materials, and analysis scripts for all experiments are provided via the Open Science Framework (OSF; https://osf.io/dt35k/). All experiments were preregistered (Experiment 1: https://osf.io/ncpvq, Experiment 2: https://osf.io/m2fjv, and Experiment 3: https://osf.io/vprxd).

Supplemental material

The supplementary material is available at qjep.sagepub.com.

Notes

1. In the context of the current research, an item refers to a test trial in the memory test, which is a binary outcome (0 = distractor chosen, 1 = target chosen).
2. A finding that there is no stochastic dependency of the retrieval of event elements in non-coherent encoding episodes (open-loop structures) would be consistent with the interplay of two assumptions: integrated binding structures and the necessity of coherent encoding episodes. However, the latter assumption is not a necessary premise for integrated binding structures. Thus, a result pattern in which there is a dependency of the retrieval of event elements in non-coherent encoding episodes but this dependency does not vary as a function of the excluded association would still be consistent with an integrated binding structure, although it would violate the assumption of the necessity of coherent encoding episodes.
3. The factor association refers to the element pair being tested in the cued recognition test. The possible associations were animal–object, animal–location, and object–location. For example, a test trial in which the cue is an animal and the target is an object and a test trial in which the cue is an object and the target is an animal both test the association animal–object.
4. The Bayesian information criterion (BIC) for a given model M can be computed as $\mathrm{BIC}(M) = k\log(n) - 2\log(L)$, where k is the number of free parameters of the model, n is the number of observations, and L is the maximum likelihood of the model. Given two models M_0 and M_1, one can approximate the Bayes factor in favour of M_1 as $\mathrm{BF}_{10} = \exp((\mathrm{BIC}_0 - \mathrm{BIC}_1)/2)$. This BIC approximation can be considered to assume the unit information prior, which contains as much information as, on average, a single observation (see Raftery, 1995; Wagenmakers, 2007).
5. In the preregistration, we noted that we would draw item parameters from a standard normal distribution. However, estimating the parameters from the data allows to better account for differences in memory performance between conditions.
6. While we initially planned to compute one-tailed p values, we switched to two-tailed p values due to the (unexpected) occurrence of negative dependencies in the data, which warranted testing.
7. In the preregistration, we mentioned that we would calculate p values based on the mean of the difference values. However, this approach may lead to a too liberal criterion because it uses more item information than is available in the empirical data. Thus, we used the distribution of the individual difference values for computing p values instead.
8. We excluded participants with time lags larger than 1 min between screens in the learning or test phase except if this occurred only once and the lag was less than 5 min. For lags between the learning and test phase, we applied a less restrictive criterion because this part of the experiment also contained instructions. Regarding lags between the learning and test phase, we excluded participants with time lags larger than 3 min except if this occurred only once and the lag was less than 7 min. Larger time lags between screens are indicative of participants interrupting the experiment.
9. One participant did not give any remember responses and four participants did not give any know responses. These participants were excluded from the respective analyses, resulting in $n = 148$ for the analysis for remember responses and $n = 145$ for the analysis for know responses.
10. A value of 0 indicates no difference in odds between groups. Positive values indicate higher odds and negative values indicate lower odds in the first group than in the second group.

11. The number of events per condition was considered in the power analysis. Thus, given the suggested sample size, the modelling approach is sufficiently robust to the lower number of events per condition.
12. The loop conditions and associations were coerced into a common factor level because they differed between the animacy and non-animacy condition, and thus needed to be "equated" to jointly include them in the generalised linear mixed model.
13. Note that condition OL-ol (open loop with association object–location excluded) in Experiment 1 also yielded a significant positive dependency but was not included in the post hoc analysis of presentation order regarding animacy.

References

Alvarez, P., & Squire, L. R. (1994). Memory consolidation and the medial temporal lobe: A simple network model. *Proceedings of the National Academy of Sciences*, *91*(15), 7041–7045. https://doi.org/10.1073/pnas.91.15.7041

Andermane, N., Joensen, B. H., & Horner, A. J. (2021). Forgetting across a hierarchy of episodic representations. *Current Opinion in Neurobiology*, *67*, 50–57. https://doi.org/10.1016/j.conb.2020.08.004

Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(5), 1063–1087. https://doi.org/10.1037/0278-7393.20.5.1063

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

Armitage, J., & Eerola, T. (2020). Reaction time data in music cognition: Comparison of pilot data from lab, crowdsourced, and convenience web samples. *Frontiers in Psychology*, *10*, Article 2883. https://doi.org/10.3389/fpsyg.2019.02883

Arnold, N. R., Heck, D. W., Bröder, A., Meiser, T., & Boywitt, C. D. (2019). Testing hypotheses about binding in context memory with a hierarchical multinomial modeling approach: A preregistered study. *Experimental Psychology*, *66*(3), 239–251. https://doi.org/10.1027/1618-3169/a000442

Aust, F., & Barth, M. (2020). *papaja: Prepare reproducible APA journal articles with R Markdown* (R package version 0.1.0.9997). https://github.com/crsh/papaja

Backus, A. R., Bosch, S. E., Ekman, M., Grabovetsky, A. V., & Doeller, C. F. (2016). Mnemonic convergence in the human hippocampus. *Nature Communications*, *7*(1), Article 11991. https://doi.org/10.1038/ncomms11991

Balaban, H., Assaf, D., Arad Meir, M., & Luria, R. (2019). Different features of real-world objects are represented in a dependent manner in long-term memory. *Journal of Experimental Psychology: General*, *149*(7), 1275–1293. https://doi.org/10.1037/xge0000716

Bartneck, C., Duenser, A., Moltchanova, E., & Zawieska, K. (2015). Comparing the similarity of responses received from studies in Amazon's Mechanical Turk to studies conducted online and with direct recruitment. *PLOS ONE*, *10*(4), Article e0121595. https://doi.org/10.1371/journal.pone.0121595

Barton, K. (2020). *MuMIn: Multi-model inference* (R package version 1.43.17). https://cran.r-project.org/package=MuMIn

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores*. Addison-Wesley.

Bonin, P., Gelin, M., & Bugaiska, A. (2014). Animates are better remembered than inanimates: Further evidence from word and picture stimuli. *Memory & Cognition*, *42*(3), 370–382. https://doi.org/10.3758/s13421-013-0368-8

Bonin, P., Gelin, M., Laroche, B., Méot, A., & Bugaiska, A. (2015). The "how" of animacy effects in episodic memory. *Experimental Psychology*, *62*(6), 371–384. https://doi.org/10.1027/1618-3169/a000308

Boywitt, C. D., & Meiser, T. (2012a). Bound context features are integrated at encoding. *Quarterly Journal of Experimental Psychology*, *65*(8), 1484–1501. https://doi.org/10.1080/17470218.2012.656668

Boywitt, C. D., & Meiser, T. (2012b). The role of attention for context-context binding of intrinsic and extrinsic features. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(4), 1099–1107. https://doi.org/10.1037/a0026988

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2013). Real-world objects are not represented as bound units: Independent forgetting of different object details from visual memory. *Journal of Experimental Psychology: General*, *142*(3), 791–808. https://doi.org/10.1037/a0029649

Bridges, D., Pitiot, A., MacAskill, M. R., & Peirce, J. W. (2020). The timing mega-study: Comparing a range of experiment generators, both lab-based and online. *PeerJ*, *8*, Article e9414. https://doi.org/10.7717/peerj.9414

Bröder, A. (2009). Semantically clustered words are stored with integrated context: Validating a measurement model for source memory, storage, and retrieval in free recall. *Zeitschrift Für Psychologie/Journal of Psychology*, *217*(3), 136–148. https://doi.org/10.1027/0044-3409.217.3.136

Cabeza, R. (2006). Prefrontal and medial temporal lobe contributions to relational memory in young and older adults. In H. Zimmer, A. Mecklinger & U. Lindenberger (Eds.), *Handbook of binding and memory: Perspectives from cognitive neuroscience* (pp. 595–626). Oxford University Press.

Cai, D. J., Aharoni, D., Shuman, T., Shobe, J., Biane, J., Song, W., Wei, B., Veshkini, M., La-Vu, M., Lou, J., Flores, S. E., Kim, I., Sano, Y., Zhou, M., Baumgaertel, K., Lavi, A., Kamata, M., Tuszynski, M., Mayford, M., & Silva, A. J. (2016). A shared neural ensemble links distinct contextual memories encoded close in time. *Nature*, *534*(7605), 115–118. https://doi.org/10.1038/nature17955

Chalmers, P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, *48*(6), 1–29. https://doi.org/10.18637/jss.v048.i06

Chalmers, P. (2020). *SimDesign: Structure for organizing Monte Carlo simulation designs* (R package version 2.0.1). https://cran.r-project.org/package=SimDesign

Chen, W.-H., & Thissen, D. (1997). Local dependence indexes for item pairs using item response theory. *Journal of*

Educational and Behavioral Statistics, *22*(3), 265–289. https://doi.org/10.3102/10769986022003265

Clifford, S., & Jerit, J. (2014). Is there a cost to convenience? An experimental comparison of data quality in laboratory and online studies. *Journal of Experimental Political Science*, *1*(2), 120–131. https://doi.org/10.1017/xps.2014.5

Cohen, N. J., & Eichenbaum, H. (1993). *Memory, amnesia, and the hippocampal system*. MIT Press.

Damasio, A. R. (1989). The brain binds entities and events by multiregional activation from convergence zones. *Neural Computation*, *1*(1), 123–132. https://doi.org/10.1162/neco.1989.1.1.123

Dandurand, F., Shultz, T. R., & Onishi, K. H. (2008). Comparing online and lab methods in a problem-solving experiment. *Behavior Research Methods*, *40*(2), 428–434. https://doi.org/10.3758/brm.40.2.428

Davachi, L., Mitchell, J. P., & Wagner, A. D. (2003). Multiple routes to memory: Distinct medial temporal lobe processes build item and source memories. *Proceedings of the National Academy of Sciences*, *100*(4), 2157–2162. https://doi.org/10.1073/pnas.0337195100

Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their application*. Cambridge University Press.

de Ayala, R. J. (2009). *The theory and practice of item response theory*. Guilford Press.

de Leeuw, J. R., & Motz, B. A. (2016). Psychophysics in a web browser? Comparing response times collected with JavaScript and Psychophysics Toolbox in a visual search task. *Behavior Research Methods*, *48*(1), 1–12. https://doi.org/10.3758/s13428-015-0567-2

DeYoung, C. M., & Serra, M. J. (2021). Judgments of learning reflect the animacy advantage for memory, but not beliefs about the effect. *Metacognition and Learning*, *16*, 711–747. https://doi.org/10.1007/s11409-021-09264-w

Eichenbaum, H. (1999). The hippocampus and mechanisms of declarative memory. *Behavioural Brain Research*, *103*(2), 123–133. https://doi.org/10.1016/S0166-4328(99)00044-3

Eichenbaum, H., & Cohen, N. J. (1988). Representation in the hippocampus: What do hippocampal neurons code? *Trends in Neurosciences*, *11*(6), 244–248. https://doi.org/10.1016/0166-2236(88)90100-2

Eichenbaum, H., & Cohen, N. J. (2001). *From conditioning to conscious recollection: Memory systems of the brain*. Oxford University Press.

Eichenbaum, H., Yonelinas, A. P., & Ranganath, C. (2007). The medial temporal lobe and recognition memory. *Annual Review of Neuroscience*, *30*, 123–152. https://doi.org/10.1146/annurev.neuro.30.051606.094328

Gardiner, J. M. (1988). Functional aspects of recollective experience. *Memory & Cognition*, *16*(4), 309–313. https://doi.org/10.3758/BF03197041

Geraci, L., McCabe, D. P., & Guillory, J. J. (2009). On interpreting the relationship between remember-know judgments and confidence: The role of instructions. *Consciousness and Cognition*, *18*(3), 701–709. https://doi.org/10.1016/j.concog.2009.04.010

Gibbons, R. D., & Hedeker, D. R. (1992). Full-information item bi-factor analysis. *Psychometrika*, *57*(3), 423–436. https://doi.org/10.1007/bf02295430

Glas, C. A. W., Wainer, H., & Bradlow, E. T. (2000). MML and EAP estimation in testlet-based adaptive testing. In W. J. van der Linden & C. A. W. Glas (Eds.), *Computerized adaptive testing: Theory and practice* (pp. 271–287). Kluwer.

Goldstein, H. (2011). *Multilevel statistical models* (4th ed.). Wiley.

Hauser, D. J., & Schwarz, N. (2016). Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior Research Methods*, *48*(1), 400–407. https://doi.org/10.3758/s13428-015-0578-z

Hayes, S. M., Ryan, L., Schnyer, D. M., & Nadel, L. (2004). An fMRI study of episodic memory: Retrieval of object, spatial, and temporal information. *Behavioral Neuroscience*, *118*(5), 885–896. https://doi.org/10.1037/0735-7044.118.5.885

Hayman, C. G., & Tulving, E. (1989). Contingent dissociation between recognition and fragment completion: The method of triangulation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(2), 228–240. https://doi.org/10.1037/0278-7393.15.2.228

Healy, M. J., & Caudell, T. P. (2019). Episodic memory: A hierarchy of spatiotemporal concepts. *Neural Networks*, *120*, 40–57. https://doi.org/10.1016/j.neunet.2019.09.021

Henninger, F., Shevchenko, Y., Mertens, U., Kieslich, P. J., & Hilbig, B. E. (2020). *lab.js: A free, open, online study builder*. https://doi.org/10.5281/zenodo.597045

Hicks, J. L., & Starns, J. J. (2004). Retrieval-induced forgetting occurs in tests of item recognition. *Psychonomic Bulletin & Review*, *11*(1), 125–130. https://doi.org/10.3758/BF03206471

Hicks, J. L., & Starns, J. J. (2016). Successful cuing of gender source memory does not improve location source memory. *Memory & Cognition*, *44*(4), 650–659. https://doi.org/10.3758/s13421-016-0586-y

Hilbig, B. E. (2016). Reaction time effects in lab- versus web-based research: Experimental evidence. *Behavior Research Methods*, *48*(4), 1718–1724. https://doi.org/10.3758/s13428-015-0678-9

Hintzman, D. L. (1972). On testing the independence of associations. *Psychological Review*, *79*(3), 261–264. https://doi.org/10.1037/h0032684

Hintzman, D. L. (1980). Simpson's paradox and the analysis of memory retrieval. *Psychological Review*, *87*(4), 398–410. https://doi.org/10.1037/0033-295x.87.4.398

Hoffman, L., & Rovine, M. J. (2007). Multilevel models for the experimental psychologist: Foundations and illustrative examples. *Behavior Research Methods*, *39*(1), 101–117. https://doi.org/10.3758/BF03192848

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*(2), 65–70.

Hommel, B. (1998). Event files: Evidence for automatic integration of stimulus-response episodes. *Visual Cognition*, *5*(1–2), 183–216. https://doi.org/10.1080/713756773

Hommel, B. (2004). Event files: Feature binding in and across perception and action. *Trends in Cognitive Sciences*, *8*(11), 494–500. https://doi.org/10.1016/j.tics.2004.08.007

Hommel, B. (2009). Action control according to TEC (theory of event coding). *Psychological Research*, *73*(4), 512–526. https://doi.org/10.1007/s00426-009-0234-2

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The Theory of Event Coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences*, *24*(5), 849–878. https://doi.org/10.1017/s0140525x01000103

Horner, A. J., Bisby, J. A., Bush, D., Lin, W.-J., & Burgess, N. (2015). Evidence for holistic episodic recollection via hippocampal pattern completion. *Nature Communications*, *6*(1), Article 7462. https://doi.org/10.1038/ncomms8462

Horner, A. J., & Burgess, N. (2013). The associative structure of memory for multi-element events. *Journal of Experimental Psychology: General*, *142*(4), 1370–1383. https://doi.org/10.1037/a0033626

Horner, A. J., & Burgess, N. (2014). Pattern completion in multielement event engrams. *Current Biology*, *24*(9), 988–992. https://doi.org/10.1016/j.cub.2014.03.012

Hunt, R. R., & Einstein, G. O. (1981). Relational and item-specific information in memory. *Journal of Verbal Learning and Verbal Behavior*, *20*(5), 497–514. https://doi.org/10.1016/S0022-5371(81)90138-9

James, E., Ong, G., Henderson, L., & Horner, A. J. (2020). Make or break it: Boundary conditions for integrating multiple elements in episodic memory. *Royal Society Open Science*, *7*(9), Article 200431. https://doi.org/10.1098/rsos.200431

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Clarendon Press.

Joensen, B. H., Gaskell, M. G., & Horner, A. J. (2020). United we fall: All-or-none forgetting of complex episodic events. *Journal of Experimental Psychology: General*, *149*(2), 230–248. https://doi.org/10.1037/xge0000648

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*(1), 3–28. https://doi.org/10.1037//0033-2909.114.1.3

Kazanas, S. A., Altarriba, J., & O'Brien, E. G. (2020). Paired-associate learning, animacy, and imageability effects in the survival advantage. *Memory & Cognition*, *48*(2), 244–255. https://doi.org/10.3758/s13421-019-01007-2

Kumaran, D., & Ludwig, H. (2013). Transitivity performance, relational hierarchy knowledge and awareness: Results of an instructional framing manipulation. *Hippocampus*, *23*(12), 1259–1268. https://doi.org/10.1002/hipo.22163

Kumaran, D., & McClelland, J. L. (2012). Generalization through the recurrent interaction of episodic memories: A model of the hippocampal system. *Psychological Review*, *119*(3), 573–616. https://doi.org/10.1037/a0028681

Lange, K., Kühn, S., & Filevich, E. (2015). "Just another tool for online studies" (JATOS): An easy solution for setup and management of web servers supporting online studies. *PLOS ONE*, *10*(6), Article e0130834. https://doi.org/10.1371/journal.pone.0130834

Laurino, J., & Kaczer, L. (2019). Animacy as a memory enhancer during novel word learning: Evidence from orthographic and semantic memory tasks. *Memory*, *27*(6), 820–828. https://doi.org/10.1080/09658211.2019.1572195

Lazarsfeld, P. F., & Henry, N. W. (1968). *Latent structure analysis*. Houghton Mifflin.

Lenth, R. (2020). *emmeans: Estimated marginal means, aka least-squares means* (R package version 1.4.7). https://cran.r-project.org/package=emmeans

Li, P., Jia, X., Li, X., & Li, W. (2016). The effect of animacy on metamemory. *Memory & Cognition*, *44*(5), 696–705. https://doi.org/10.3758/s13421-016-0598-7

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Lawrence Erlbaum.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Addison-Wesley.

Madan, C. R. (2021). Exploring word memorability: How well do different word properties explain item free-recall probability? *Psychonomic Bulletin & Review*, *28*(2), 583–595. https://doi.org/10.3758/s13423-020-01820-w

Marr, D. (1971). Simple memory: A theory for archicortex. *Philosophical Transactions of the Royal Society of London. B: Biological Sciences*, *262*(841), 23–81. https://doi.org/10.1098/rstb.1971.0078

McClelland, J. L., O'Reilly, R. C., & McNaughton, B. L. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*(3), 419–457. https://doi.org/10.1037/0033-295x.102.3.419

Meiser, T., & Bröder, A. (2002). Memory for multidimensional source information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(1), 116–137. https://doi.org/10.1037//0278-7393.28.1.116

Meiser, T., Sattler, C., & Weißer, K. (2008). Binding of multidimensional context information as a distinctive characteristic of remember judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(1), 32–49. https://doi.org/10.1037/0278-7393.34.1.32

Migo, E. M., Mayes, A. R., & Montaldi, D. (2012). Measuring recollection and familiarity: Improving the remember/know procedure. *Consciousness and Cognition*, *21*(3), 1435–1455. https://doi.org/10.1016/j.concog.2012.04.014

Moeller, B., Pfister, R., Kunde, W., & Frings, C. (2019). Selective binding of stimulus, response, and effect features. *Psychonomic Bulletin & Review*, *26*(5), 1627–1632. https://doi.org/10.3758/s13423-019-01646-1

Moll, M., & Miikkulainen, R. (1997). Convergence-zone episodic memory: Analysis and simulations. *Neural Networks*, *10*(6), 1017–1036. https://doi.org/10.1016/S0893-6080(97)00016-6

Morton, J., Hammersley, R. H., & Bekerian, D. A. (1985). Headed records: A model for memory and its failures. *Cognition*, *20*(1), 1–23. https://doi.org/10.1016/0010-0277(85)90002-2

Morton, N. W., Schlichting, M. L., & Preston, A. R. (2020). Representations of common event structure in medial temporal lobe and frontoparietal cortex support efficient inference. *Proceedings of the National Academy of Sciences*, *117*(47), 29338–29345. https://doi.org/10.1073/pnas.1912338117

Nairne, J. S., Pandeirada, J. N. S., & Thompson, S. R. (2008). Adaptive memory: The comparative value of survival processing. *Psychological Science*, *19*(2), 176–180. https://doi.org/10.1111/j.1467-9280.2008.02064.x

Nairne, J. S., Thompson, S. R., & Pandeirada, J. N. S. (2007). Adaptive memory: Survival processing enhances retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(2), 263–273. https://doi.org/10.1037/0278-7393.33.2.263

Nairne, J. S., VanArsdall, J. E., & Cogdill, M. (2017). Remembering the living: Episodic memory is tuned to animacy. *Current Directions in Psychological Science*, *26*(1), 22–27. https://doi.org/10.1177/0963721416667711

Nairne, J. S., VanArsdall, J. E., Pandeirada, J. N. S., Cogdill, M., & LeBreton, J. M. (2013). Adaptive memory: The mnemonic value of animacy. *Psychological Science*, *24*(10), 2099–2105. https://doi.org/10.1177/0956797613480803

Nakagawa, S., Johnson, P. C. D., & Schielzeth, H. (2017). The coefficient of determination R^2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of the Royal Society Interface*, *14*(134), Article 20170213. https://doi.org/10.1098/rsif.2017.0213

Ngo, C. T., Horner, A. J., Newcombe, N. S., & Olson, I. R. (2019). Development of holistic episodic recollection. *Psychological Science*, *30*(12), 1696–1706. https://doi.org/10.1177/0956797619879441

Popp, E. Y., & Serra, M. J. (2016). Adaptive memory: Animacy enhances free recall but impairs cued recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(2), 186–201. https://doi.org/10.1037/xlm0000174

Preston, A. R., Shrager, Y., Dudukovic, N. M., & Gabrieli, J. D. E. (2004). Hippocampal contribution to the novel use of relational information in declarative memory. *Hippocampus*, *14*(2), 148–152. https://doi.org/10.1002/hipo.20009

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, 25, 111–163. https://doi.org/10.2307/271063

Reinitz, M. T., Lammers, W. J., & Cochran, B. P. (1992). Memory-conjunction errors: Miscombination of stored stimulus features can produce illusions of memory. *Memory & Cognition*, *20*(1), 1–11. https://doi.org/10.3758/BF03208247

Robin, J. (2018). Spatial scaffold effects in event memory and imagination. *WIREs Cognitive Science*, *9*(4), Article e1462. https://doi.org/10.1002/wcs.1462

Robitzsch, A. (2020). *sirt: Supplementary item response theory models* (R package version 3.9-4). https://cran.r-project.org/package=sirt

Rubin, D. C., & Umanath, S. (2015). Event memory: A theory of memory for laboratory, autobiographical, and fictional events. *Psychological Review*, *122*(1), 1–23. https://doi.org/10.1037/a0037907

Schreiner, M. R., & Meiser, T. (2022). Measuring binding effects in event-based episodic representations. *Behavior Research Methods*. https://doi.org/10.3758/s13428-021-01769-1

Shohamy, D., & Wagner, A. D. (2008). Integrating memories in the human brain: Hippocampal-midbrain encoding of overlapping events. *Neuron*, *60*(2), 378–389. https://doi.org/10.1016/j.neuron.2008.09.023

Simpson, E. H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society*, *13*(2), 238–241. https://doi.org/10.1111/j.2517-6161.1951.tb00088.x

Squire, L., & Zola-Morgan, S. (1991). The medial temporal lobe memory system. *Science*, *253*(5026), 1380–1386. https://doi.org/10.1126/science.1896849

Starns, J. J., & Hicks, J. L. (2005). Source dimensions are retrieved independently in multidimensional monitoring tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(6), 1213–1220. https://doi.org/10.1037/0278-7393.31.6.1213

Starns, J. J., & Hicks, J. L. (2008). Context attributes in memory are bound to item information, but not to one another. *Psychonomic Bulletin & Review*, *15*(2), 309–314. https://doi.org/10.3758/PBR.15.2.309

SurveyCircle. (2021). *Research website SurveyCircle*. https://www.surveycircle.com

Trinkler, I., King, J., Spiers, H. J., & Burgess, N. (2006). Part or parcel? Contextual binding of events in episodic memory. In H. Zimmer, A. Mecklinger & U. Lindenberger (Eds.), *Handbook of binding and memory: Perspectives from cognitive neuroscience* (pp. 53–83). Oxford University Press.

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory* (pp. 381–403). Academic Press.

Tulving, E. (1983). *Elements of episodic memory*. Oxford University Press.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology/Psychologie Canadienne*, *26*(1), 1–12. https://doi.org/10.1037/h0080017

Utochkin, I. S., & Brady, T. F. (2020). Independent storage of different features of real-world objects in long-term memory. *Journal of Experimental Psychology: General*, *149*(3), 530–549. https://doi.org/10.1037/xge0000664

VanArsdall, J. E., Nairne, J. S., Pandeirada, J. N. S., & Blunt, J. R. (2013). Adaptive memory: Animacy processing produces mnemonic advantages. *Experimental Psychology*, *60*(3), 172–178. https://doi.org/10.1027/1618-3169/a000186

VanArsdall, J. E., Nairne, J. S., Pandeirada, J. N. S., & Cogdill, M. (2015). Adaptive memory: Animacy effects persist in paired-associate learning. *Memory*, *23*(5), 657–663. https://doi.org/10.1080/09658211.2014.916304

Vogt, V., & Bröder, A. (2007). Independent retrieval of source dimensions: An extension of results by Starns and Hicks (2005) and a comment on the ACSIM measure. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(2), 443–450. https://doi.org/10.1037/0278-7393.33.2.443

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, *14*(5), 779–804. https://doi.org/10.3758/BF03194105

Wainer, H., & Wang, X. (2000). Using a new statistical model for testlets to score TOEFL. *Journal of Educational Measurement*, *37*(3), 203–220. https://doi.org/10.1111/j.1745-3984.2000.tb01083.x

Wang, X., Bradlow, E. T., & Wainer, H. (2002). A general Bayesian model for testlets: Theory and applications. *Applied Psychological Measurement*, *26*(1), 109–128. https://doi.org/10.1177/0146621602026001007

Yen, W. M. (1984). Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Applied Psychological Measurement*, *8*(2), 125–145. https://doi.org/10.1177/014662168400800201

Yen, W. M. (1993). Scaling performance assessments: Strategies for managing local item dependence. *Journal of Educational Measurement*, *30*(3), 187–213. https://doi.org/10.1111/j.1745-3984.1993.tb00423.x

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, *46*(3), 441–517. https://doi.org/10.1006/jmla.2002.2864

Yule, G. U. (1912). On the methods of measuring association between two attributes. *Journal of the Royal Statistical Society*, *75*(6), 579–652. https://doi.org/10.2307/2340126

Zeithamova, D., Schlichting, M. L., & Preston, A. R. (2012). The hippocampus and inferential reasoning: Building memories to navigate future decisions. *Frontiers in Human Neuroscience*, *6*, Article 70. https://doi.org/10.3389/fnhum.2012.00070

Zotow, E., Bisby, J. A., & Burgess, N. (2020). Behavioral evidence for pattern separation in human episodic memory. *Learning & Memory*, *27*(8), 301–309. https://doi.org/10.1101/lm.051821.120

Appendix A

Bifactor IRT Model

The bifactor IRT model we used extends the model in (1) by including additional latent traits λ for each event:

$$P(u_{ij} = 1) = \gamma_j + (1 - \gamma_j) \frac{e^{\alpha_j(\theta_i - \beta_j) - \alpha_{t(j)}\lambda_{it(j)}}}{1 + e^{\alpha_j(\theta_i - \beta_j) - \alpha_{t(j)}\lambda_{it(j)}}} \tag{A1}$$

where λ is the event-specific trait of person *i* for event *t(j)* to which item *j* belongs. Thus, each item loads on the general latent person trait and on an additional event-specific latent trait. Applying the same restrictions as in (2) reduces the model to:

$$P(u_{ij} = 1) = \frac{1}{6} + \frac{5}{6} \frac{e^{\theta_i - \beta_j - \lambda_{it(j)}}}{1 + e^{\theta_i - \beta_j - \lambda_{it(j)}}} \tag{A2}$$

The event-specific latent traits are mutually independent and independent to the general latent person trait θ. They exert their influence via their variances, with higher variances indicating a larger event-specific effect.

Appendix B

Dependency Analysis Using the Approach by Horner and Burgess

We also analysed the dependency of the retrieval of event elements using the approach outlined in Horner and Burgess (2013), using code provided by James et al. (2020). Because Horner and Burgess (2013) tested associations in both directions, resulting in six test trials per event, the approach had to be adapted to fit the current procedure, which tested associations in only one direction, resulting in three test trials per event. While Horner and Burgess (2013) identified test pairs based on a common cue or target element, we identified test pairs based on the occurrence of a common element in the tested associations. For example, for the animacy conditions this resulted in three test pairs: the trials testing animal – object and animal – location, the trials testing animal – object and object – location, and the trials testing animal – location and object – location. The approach yields participant-specific dependency estimates for each condition. We tested for the presence of dependency in each condition using one-sample *t*-tests against 0. We compared dependency across conditions using linear mixed models and performing planned comparisons (one-tailed testing). For the linear mixed models we report Bayes factors in favour of an effect obtained using BIC approximation (Raftery, 1995; Wagenmakers, 2007) and the change in marginal R^2 (Nakagawa et al., 2017). In Experiments 2 and 3, loop conditions were coerced for the linear mixed model as done for the memory performance analysis.

In Experiment 1 there were significant positive dependencies in all conditions ($M_{\text{CL}} = 0.02$, $t(148) = 3.41$, $p = .001$, $d = 0.28$; $M_{\text{OL-ao}} = 0.01$, $t(148) = 3.01$, $p = .003$, $d = 0.25$; $M_{\text{OL-al}} = 0.01$, $t(148) = 3.11$, $p = .002$, $d = 0.25$; $M_{\text{OL-ol}} = 0.03$, $t(148) = 5.27$, $p < .001$, $d = 0.43$). There was strong evidence against an effect of condition according to the linear mixed model analysis ($\text{BF}_{10} < 0.001$, $R^2_{\text{change}} = .01$). Planned comparisons revealed that the dependency in condition CL was not significantly higher compared to the open-loop conditions ($t(444) = -0.37$, $p = .64$). However, the dependency in condition OL-ol was significantly higher compared to conditions OL-ao and OL-al ($t(444) = 2.65$, $p = .004$).

These results are largely congruent with the ones from the main dependency analysis. They support Hypothesis 1, which stated that there is a stochastic dependency of the retrieval of event elements and Hypothesis 4, which stated that dependency varies as a function of the excluded association in non-coherent encoding episodes. The results are thus in favour of a hierarchical binding structure. Contrary to the main dependency analysis, the significant yet weak dependencies in conditions OL-ao and OL-al partially support Hypothesis 3, which stated that dependency is reduced but not eliminated in non-coherent encoding episodes.

In Experiment 2, there were significant positive dependencies in conditions OL-ao ($M = 0.01$, $t(212) = 2.29$, $p = .02$, $d = 0.16$) and OL-ol ($M = 0.01$, $t(212) = 2.68$, $p = .008$, $d = 0.18$) in the animacy condition, but no significant dependencies in conditions CL ($M = 0.00$, $t(212) = 0.56$, $p = .58$, $d = 0.04$) and OL-al ($M = 0.00$, $t(212) = 0.88$, $p = .38$, $d = 0.06$). There were no significant dependencies in the non-animacy condition ($M_{\text{CL}} = 0.01$, $t(212) = 1.44$, $p = .58$, $d = 0.10$; $M_{\text{OL-o}_{\text{tr}}\text{o}_{\text{to}}} = 0.00$, $t(212) = -0.02$, $p = .98$, $d = 0.00$; $M_{\text{OL-o}_{\text{tr}}\text{l}} = 0.00$, $t(212) = 1.00$, $p = .32$, $d = 0.07$; $M_{\text{OL-o}_{\text{to}}\text{l}} = 0.00$, $t(212) = 0.70$, $p = .48$, $d = 0.05$). There was strong evidence against main effects of animacy condition ($\text{BF}_{10} < 0.001$, $R^2_{\text{change}} = .001$) and loop condition ($\text{BF}_{10} < 0.001$, $R^2_{\text{change}} < .001$) and against an interaction ($\text{BF}_{10} < 0.001$, $R^2_{\text{change}} = .002$) according to the linear mixed model analysis. Planned comparisons revealed that the dependency in condition CL was not significantly higher compared to the open-loop conditions in the animacy condition ($t(1484) = -1.09$, $p = .86$) and in the non-animacy condition ($t(1484) = 0.86$, $p = .19$). The dependency in condition OL-ol was not significantly higher compared to conditions OL-ao and OL-al in the animacy condition ($t(1484) = 0.97$, $p = .17$). The dependency in condition OL-o_{to}l was not significantly higher compared to conditions OL-$\text{o}_{\text{tr}}\text{o}_{\text{to}}$ and OL-o_{tr}l in the non-animacy condition ($t(1484) = 0.23$, $p = .41$).

These results are largely congruent with the ones from the main dependency analysis. While, contrary to the main dependency analysis, the dependency in condition OL-ao and OL-ol in the animacy condition reached significance, the effects were weak and the dependency in condition OL-ol was not significantly larger than in the other open-loop conditions. Like the results of the main dependency analysis, the result can neither clearly distinguish between an integrated and a hierarchical binding structure, nor do they provide evidence for a special role of animacy.

In Experiment 3, there was a significant positive dependency in condition CL in the animacy condition (M = 0.02, $t(130)$ = 2.99, p = .003, d = 0.26) but no significant dependencies in the open-loop conditions ($M_{\text{OL-ao}}$ = 0.01, $t(130)$ = 1.43, p = .15, d = 0.13; $M_{\text{OL-al}}$ = 0.00, $t(130)$ = 0.77, p = .44, d = 0.07; $M_{\text{OL-ol}}$ = 0.01, $t(130)$ = 1.48, p = .14, d = 0.13). There were no significant dependencies in the non-animacy condition (M_{CL} = 0.01, $t(121)$ = 0.88, p = .38, d = 0.08; $M_{\text{OL-o}_{\text{tr}}\text{o}_{\text{to}}}$ = 0.00, $t(121)$ = -0.85, p = .40, d = -0.08; $M_{\text{OL-o}_{\text{tr}}\text{l}}$ = -0.01, $t(121)$ = -1.35, p = .18, d = -0.12; $M_{\text{OL-o}_{\text{to}}\text{l}}$ = 0.00, $t(121)$ = 0.25, p = .80, d = 0.02). There was strong evidence against main effects of animacy condition (BF_{10} = 0.005, R^2_{change} = .007) and loop condition ($\text{BF}_{10} < 0.001$, R^2_{change} = .005) and against an interaction ($\text{BF}_{10} < 0.001$, $R^2_{\text{change}} < .001$) according to the linear mixed model analysis. Planned comparisons revealed that the dependency in condition CL was significantly higher compared to the open-loop conditions in the animacy condition ($t(753)$ = 1.81, p = .04) but not in the non-animacy condition ($t(753)$ = 1.38, p = .08). The dependency in condition OL-ol was not significantly higher compared to conditions OL-ao and OL-al in the animacy condition ($t(753)$ = 0.26, p = .40). The dependency in condition OL-o_{to}l was not significantly higher compared to conditions OL-$\text{o}_{\text{tr}}\text{o}_{\text{to}}$ and OL-o_{tr}l in the non-animacy condition ($t(753)$ = 1.02, p = .15).

These results are largely congruent with the main dependency analysis except for the non-significant dependencies in the open-loop conditions in the non-animacy condition, which were significantly negative in the main dependency analysis. The results support Hypothesis 1 but do not support Hypotheses 3 and 4. They are thus in favour of an integrated binding structure. The results further support Hypothesis 5b, which stated that dependency does not vary as a function of the excluded association in non-coherent encoding episodes if events do not include an animate element, but do not support Hypothesis 5a, which stated that dependency varies as a function of the excluded association in non-coherent encoding episodes if events include an animate element. However, like the results of the main dependency analysis, the results suggest that animacy facilitates the binding of event elements if the encoding episode is coherent.

Appendix C

Power Analysis Description

Power analyses were conducted using Monte Carlo simulations. Data were drawn from the bifactor model in (A2) using 1000 replications. Differences between conditions were induced by specifying differences in the variances of event-specific latent traits. For example, a difference of 1 was considered a medium effect (cf. Glas et al., 2000; Wang et al., 2002). This value was subtracted from the event-specific trait variance of a reference condition. Latent factor variances of both the general person trait and the reference event-specific trait were based on the ones from a pilot study ($N = 27$; Experiment 1) or on the ones from Experiment 1 (Experiments 2 and 3). The pilot study was using the simultaneous encoding paradigm (Horner & Burgess, 2013, 2014) and yielded higher dependency estimates than the separated encoding paradigm (Horner et al., 2015; Horner & Burgess, 2014; see also James et al., 2020) used in the main experiments, thus leading to a higher baseline dependency in the power analysis. This makes the power analysis more conservative. Reference event-specific trait variances were 4 for the power analysis for Experiment 1 and 1 for the power analyses for Experiments 2 and 3. The power analyses required critical values to which the estimated differences in dependency could be compared to determine statistical significance. These critical values were obtained by conducting other Monte Carlo simulations. Data were again drawn from the bifactor model in (A2) using 1000 replications. However, in these simulations all event-specific traits had the same variance relative to the reference condition. Then, difference values were computed and the 5% and 95% quantiles were used as critical values in the power analyses (one-tailed testing). The procedure was thus similar to the parametric bootstrap. Power analyses were targeted at the predicted pattern of effects. Thus, for each replication it was determined whether all predicted differences were significant. Power was determined over a range of sample sizes with an increment of 10.

Appendix D

Memory Performance by Cue, Target, and Trial Type

Table D1

Mean (M) and Standard Deviation (SD) of Memory Performance by Cue, Target, and Trial Type in each Condition of Experiment 1

		Condition							
		CL		OL-ao		OL-al		OL-ol	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Cue type	animal	0.53	0.50	0.42	0.49	0.45	0.50	0.52	0.50
	object	0.57	0.50	0.53	0.50	0.46	0.50	0.42	0.49
	location	0.53	0.50	0.44	0.50	0.54	0.50	0.43	0.49
Target type	animal	0.55	0.50	0.41	0.49	0.44	0.50	0.50	0.50
	object	0.56	0.50	0.53	0.50	0.46	0.50	0.43	0.50
	location	0.55	0.50	0.45	0.50	0.55	0.50	0.43	0.50
Trial type	inference	–	–	0.33	0.47	0.36	0.48	0.35	0.48
	no inference	0.55	0.50	0.53	0.50	0.55	0.50	0.51	0.50

Note. CL = closed loop, OL-ao = open loop with association animal – object excluded, OL-al = open loop with association animal – location excluded, OL-ol = open loop with association object – location excluded.

Table D2

Mean (M) and Standard Deviation (SD) of Memory Performance by Cue, Target, and Trial Type in each Condition of Experiment 2

Animacy condition			Loop condition							
			CL		OL-ao/$o_{tr}o_{to}$		OL-al/$o_{tr}l$		OL-ol/$o_{to}l$	
			M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Animacy	Cue type	animal	0.46	0.50	0.34	0.48	0.35	0.48	0.44	0.50
		object	0.45	0.50	0.45	0.50	0.34	0.48	0.34	0.47
		location	0.46	0.50	0.35	0.48	0.42	0.49	0.35	0.48
	Target type	animal	0.45	0.50	0.36	0.48	0.34	0.47	0.45	0.50
		object	0.48	0.50	0.44	0.50	0.33	0.47	0.35	0.48
		location	0.44	0.50	0.34	0.47	0.45	0.50	0.33	0.47
	Trial type	inference	–	–	0.25	0.43	0.24	0.43	0.24	0.43
		no inference	0.46	0.50	0.44	0.50	0.44	0.50	0.45	0.50
Non-animacy	Cue type	transport	0.42	0.49	0.35	0.48	0.32	0.47	0.42	0.49
		tool	0.44	0.50	0.45	0.50	0.33	0.47	0.34	0.48
		location	0.43	0.49	0.34	0.48	0.41	0.49	0.34	0.47
	Target type	transport	0.43	0.50	0.34	0.47	0.31	0.46	0.43	0.50
		tool	0.44	0.50	0.45	0.50	0.34	0.47	0.35	0.48
		location	0.42	0.49	0.35	0.48	0.41	0.49	0.33	0.47
	Trial type	inference	–	–	0.24	0.43	0.24	0.43	0.25	0.44
		no inference	0.43	0.50	0.45	0.50	0.41	0.49	0.42	0.49

Note. CL = closed loop, OL-ao = open loop with association animal – object excluded, OL-al = open loop with association animal – location excluded, OL-ol = open loop with association object – location excluded, OL-$o_{tr}o_{to}$ = open loop with association means of transportation – tool excluded, OL-$o_{tr}l$ = open loop with association means of transportation – location excluded, OL-$o_{to}l$ = open loop with association tool – location excluded, transport = means of transportation. When two loop conditions are separated by a slash (/) the first one refers to the animacy condition and the second one refers to the non-animacy condition.

Table D3

Mean (M) and Standard Deviation (SD) of Memory Performance by Cue, Target, and Trial Type in each Condition of Experiment 3

Animacy condition			Loop condition							
			CL		OL-ao/$o_{tr}o_{to}$		OL-al/o_{tr}l		OL-ol/o_{to}l	
			M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Animacy	Cue type	animal	0.49	0.50	0.39	0.49	0.38	0.49	0.48	0.50
		object	0.53	0.50	0.50	0.50	0.39	0.49	0.38	0.49
		location	0.52	0.50	0.40	0.49	0.49	0.50	0.39	0.49
	Target type	animal	0.50	0.50	0.39	0.49	0.37	0.48	0.48	0.50
		object	0.53	0.50	0.50	0.50	0.40	0.49	0.39	0.49
		location	0.51	0.50	0.40	0.49	0.49	0.50	0.38	0.48
	Trial type	inference	–	–	0.29	0.45	0.28	0.45	0.29	0.45
		no inference	0.51	0.50	0.50	0.50	0.49	0.50	0.48	0.50
Non-animacy	Cue type	transport	0.44	0.50	0.36	0.48	0.32	0.46	0.43	0.49
		tool	0.45	0.50	0.44	0.50	0.34	0.48	0.33	0.47
		location	0.46	0.50	0.34	0.47	0.42	0.49	0.31	0.46
	Target type	transport	0.44	0.50	0.34	0.47	0.30	0.46	0.44	0.50
		tool	0.48	0.50	0.49	0.50	0.33	0.47	0.34	0.47
		location	0.43	0.50	0.32	0.47	0.44	0.50	0.29	0.45
	Trial type	inference	–	–	0.22	0.41	0.22	0.41	0.20	0.40
		no inference	0.45	0.50	0.46	0.50	0.43	0.50	0.43	0.50

Note. CL = closed loop, OL-ao = open loop with association animal – object excluded, OL-al = open loop with association animal – location excluded, OL-ol = open loop with association object – location excluded, OL-$o_{tr}o_{to}$ = open loop with association means of transportation – tool excluded, OL-o_{tr}l = open loop with association means of transportation – location excluded, OL-o_{to}l = open loop with association tool – location excluded, transport = means of transportation. When two loop conditions are separated by a slash (/) the first one refers to the animacy condition and the second one refers to the non-animacy condition.

Agency Effects on the Binding of Event Elements in Episodic Memory

Marcel R. Schreiner, Arndt Bröder, and Thorsten Meiser
Department of Psychology
School of Social Sciences
University of Mannheim

Author Note

Marcel R. Schreiner https://orcid.org/0000-0001-7719-6281, Address: L13, 15, 68161 Mannheim, Germany, E-mail: m.schreiner@uni-mannheim.de

Arndt Bröder https://orcid.org/0000-0002-2658-4791, Address: L13, 17, 68161 Mannheim, Germany, E-mail: broeder@uni-mannheim.de

Thorsten Meiser https://orcid.org/0000-0001-6004-9787, Address: L13, 15, 68161 Mannheim, Germany, E-mail: t.meiser@uni-mannheim.de

The designs, hypotheses, and analysis plans of all experiments were preregistered (Experiments 1 and 2: https://osf.io/kts8p, Experiment 3: https://osf.io/vhmt4, Experiment 4: https://osf.io/q5tme, Experiment 5: https://osf.io/g59uh). All data, materials, and analysis scripts are provided via the Open Science Framework (https://osf.io/3pkmf/).

Funding: This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - GRK 2277 “Statistical Modeling in Psychology”.

We have no conflicts of interest to disclose.

The authors made the following contributions. Marcel R. Schreiner: Conceptualisation, Methodology, Formal Analysis, Investigation, Software, Data curation, Validation, Visualization, Writing - original draft, Writing - review & editing; Arndt Bröder: Conceptualization, Supervision, Writing - review & editing; Thorsten Meiser: Conceptualization, Supervision, Writing - review & editing.

Correspondence concerning this article should be addressed to Marcel R. Schreiner, L13, 15, 68161 Mannheim, Germany. E-mail: m.schreiner@uni-mannheim.de

Abstract

Representing events in episodic memory in a coherent manner requires the events' constituent elements to be bound together. So far, only few moderators of these binding processes have been identified. Here we investigate whether the presence of an agentic element in an event facilitates binding. The results from five experiments hinted at a facilitating effect of agency on the binding of event elements. In addition, binding effects were only found when event elements were presented simultaneously, but not when they were presented sequentially pairwise, contrary to previous findings. The results suggest that the presence of an agentic element in an event may facilitate the formation of coherent memory representations and that additional processes may be required when binding event elements across temporarily divided encoding episodes. These findings add to a growing body of research regarding moderators and processes relevant for the binding of event elements in episodic memory. Explanations of these findings and directions for future research are discussed.

Keywords: episodic memory, binding, agency, statistical modeling

Word count: 13,364

Agency Effects on the Binding of Event Elements in Episodic Memory

Experienced events stored in episodic memory encompass multiple elements, such as persons, objects, locations, actions, and sensations (Tulving, 1972, 1983). To allow for coherent event representations, these elements need to be bound together in memory. Such binding processes are associated with a stochastic dependency of the retrieval of event elements, such that the successful retrieval of an event element is associated with an increased likelihood of successful retrieval of subsequent event elements (e.g., Arnold et al., 2019; Boywitt & Meiser, 2012a, 2012b; Bröder, 2009; Horner et al., 2015; Horner & Burgess, 2013, 2014; Joensen et al., 2020; Meiser & Bröder, 2002; Schreiner et al., 2022; Starns & Hicks, 2005, 2008). Despite the importance of these binding processes for the formation and retrieval of episodic memory representations, research on moderators influencing these binding processes has been scarce.

Only a small number of studies investigated moderators on the binding of event elements, encompassing aspects of stimulus presentation, event structure awareness, and animacy. James et al. (2020) identified the modality of stimulus presentation and its dimensionality as potential moderators influencing the binding of event elements, with written (rather than pictorial) stimuli and unidimenisonal (rather than multidimensional) stimulus presentation facilitating binding processes. There is also some evidence that awareness regarding the structure of an event (e.g., the number and types of elements that make up an event) is important for successful binding (Kumaran & Ludwig, 2013; Morton et al., 2020; Schreiner et al., 2022). In addition, animacy plays a major role in human memory (Nairne et al., 2013, 2017). According to the animacy effect, words representing animate entities are retrieved more likely than words representing inanimate entities (e.g., Li et al., 2016; Nairne et al., 2013; VanArsdall et al., 2015). Animate entities are living things that are capable of independent movement and can change direction without warning (Bonin et al., 2015). The animacy effect on memory has been found across a variety of test formats, including free recall (Bonin et al., 2015; Leding, 2019; Li et al., 2016; Madan, 2021; Nairne et al., 2013; Popp & Serra, 2016), recognition (Bonin et al., 2014; Leding, 2020; VanArsdall et al., 2013), and judgments of learning (DeYoung & Serra, 2021; Li et al., 2016). Results using cued recall tests have been mixed, with some studies finding the animacy effect (DeYoung & Serra, 2021; Laurino & Kaczer, 2019; VanArsdall et al., 2015) and others finding an opposite effect (Kazanas et al., 2020; Popp & Serra, 2016). Beyond enhancing memory performance, we previously found first evidence suggesting that animacy also facilitates the binding of event elements in episodic memory (Schreiner et al., 2022).

A potential explanation for the facilitating effect of animacy on the binding of event elements may be that the presence of an animate element provides a potential agent in an event. The concepts of animacy and agency are highly confounded, since animates are typically agentic. Thus, previously observed effects of animacy on binding (Schreiner et al., 2022) may actually be driven by *agency*. If this is the case, similar effects should be observed when event elements are equated regarding their animacy, but differ regarding their agency. Agency may be considered a property of animacy and may, in principle, also extend to inanimate elements (e.g., Lowder & Gordon, 2015). Animacy itself may thus be only one of several factors driving agency. Another factor may be the actual performance of an action. For example, an animal performing an action may be perceived as being more agentic than a passive animal or an animal that is the recipient of an action. Thus, agency may be a more proximate explanation for effects of animacy on binding. Agency can be defined as “acting or having the capacity to act autonomously in a given environment” (Suitner & Maass, 2016, p. 248; see also Hitlin & Elder, 2007) and is associated with concepts such as control over an action, dominance, competence, activity, and efficiency (Abele et al., 2008; Abele & Wojciszke, 2007; Bandura, 1989; Wojciszke et al., 2009). Agency plays an important role in status perception and stereotype formation (e.g., Carrier et al., 2014; Conway et al., 1996; Koch et al., 2016). Research on effects of agency in relation to memory has been scarcer. Most studies focused on agency on the participants’ side. For example, Woike et al. (1999) and Woike and Polo (2001) found the agency orientation of participants to affect the content and structure of recalled autobiographical memories. Jainta et al. (2022) found effects of agency during encoding on episodic memory, such that participants exhibited stronger hippocampal responses to expectation violations when they were actors rather than observers in the episode. Self-performed episodes were also found to be remembered better than observed ones (Hornstein & Mulligan, 2001). Huffman and Brockmole (2020) and Wen and Haggard (2018) found a bias in visual attention for objects that were under the participants' control, thus invoking a sense of agency. Stimuli over which one feels a sense of agency are also remembered better than stimuli for which this is not the case (Hon & Yeo, 2021). Regarding the agency of (external) stimuli, Walker and Keller (2019) found a processing advantage for faces with attractive, likable, and agentic traits. In addition, a major principle in the organization of object vision is a graded distinction between animate and inanimate entities in the ventral temporal cortex (an animacy continuum, Connolly et al., 2012; Sha et al., 2015; Thorat et al., 2019), to which agency is an important contributor (Haxby et al., 2020; Thorat et al., 2019). In the current research we investigate whether agency as part of the stimulus facilitates the binding of event elements in episodic memory. Considering that animacy effects in memory are commonly explained by survival-relatedness, originating from selective pressure on our ancestors (e.g., animate entities are potential prey or opponents, Nairne et al., 2007, 2008, 2013), a similar reasoning may be applied to agency. For example, agentic entities may be particularly dangerous opponents.

We investigated the role of agency in the binding of event elements in five experiments. In Experiments 1-3 event elements were presented sequentially pairwise (cf. Horner et al., 2015; Horner & Burgess, 2014), whereas they were presented simultaneously (cf. Horner & Burgess, 2013) in Experiments 4 and 5. The sequential pairwise presentation provides a very strict test of binding, because coherent memory representations need to be formed across several temporarily divided encoding episodes. Thus, binding effects are indicative of a pure form of binding in memory, because they are less likely to occur due to covariations in perceptual variables. However, given the reduced temporal contiguity of encoding episodes compared to simultaneous presentation of event elements, this form of event presentation deviates from how events are naturally experienced. Binding effects given simultaneous presentation of event elements tend to be more robust (see James et al., 2020). We found no significant dependencies of the retrieval of event elements when event elements were presented sequentially pairwise. Given simultaneous presentation of event elements however, the results hinted at a facilitating effect of agency on the binding of event elements. Experiment 2 was additionally designed to investigate whether event elements are bound in an integrated manner or in a hierarchical manner in which elements are preferentially bound to the event's agent (cf. Horner et al., 2015; Horner & Burgess, 2014; Schreiner et al., 2022), but yielded uninformative results concerning this question.

Data Availability

The data, materials, and analysis code for all experiments are provided via the Open Science Framework (https://osf.io/3pkmf/) and the designs, hypotheses, and analysis plans of all experiments were preregistered.

Experiment 1

In Experiment 1 we investigated whether the presence of an agentic element in an event facilitates binding, using a linguistic manipulation of agency. In sentences containing interpersonal action verbs (e.g., *hit)*, the agent tends to be the grammatical subject, whereas the patient of an action tends to be the grammatical object (Kasof & Lee, 1993). Consequently, grammatical subjects are perceived as being more agent-like than grammatical objects (Kako, 2006) and, for action verbs, greater causal weight is given to the agent than to the patient (Brown & Fish, 1983; Kassin & Lowe, 1979). For example, animacy tends to be a strong predictor of subject assignment (Prentice, 1967) and animate referents are usually agentic. In addition, there is an influence of transitivity. Fausey and Boroditsky (2010) found linguistic framing to influence participants' judgments of blame and financial liability. People who read transitive agentive frames (e.g., *Timberlake ripped the costume.*) allocated higher blame and financial liability than people who read intransitive non-agentive frames (e.g., *The costume ripped.*). A potential mechanism for these effects may be conceptual accessibility, which describes the ease of activation or retrieval of mental representations of a potential referent (Bock & Warren, 1985). Both animate and agentic referents are more conceptually accessible than inanimate or patient referents (Gleitman et al., 2007; Prat-Sala & Branigan, 2000; Rissman et al., 2019). Another explanation may be that agents (and grammatical subjects) are more salient than patients (and grammatical objects), particularly in third-person interpersonal action sentences (Kasof & Lee, 1993; Prat-Sala & Branigan, 2000).

We thus created agentic event elements by placing them as grammatical subjects in transitive active sentences (e.g., *The bicycle grabs the hammer.*), whereas the non-agentic elements were placed as the grammatical objects. If the sentence contained only non-agentic elements we used passive sentences (e.g., *The hammer and the bicycle are being grabbed.*), in which the grammatical subject is not the agent of the event (Kako, 2006). The use of active verb forms in the agency condition is an additional component that should increase perceived agency, whereas the use of passive verb forms should diminish it (see e.g., Henley et al., 1995; see also Halicki et al., 2021). We used only inanimate stimuli to experimentally isolate potential effects of agency from effects of animacy (cf. Schreiner et al., 2022). We used the separated encoding paradigm (Horner et al., 2015; Horner & Burgess, 2014; see also Schreiner et al., 2022), in which each pairwise association in an event is presented separately during encoding, interleaved with learning trials from other events. Previous studies found comparable dependencies using separated encoding or simultaneous encoding, in which all event elements are presented in a single learning trial (Bisby et al., 2018; Horner & Burgess, 2014; but see James et al., 2020 for boundary conditions). Since our events consisted of three elements, participants were presented three sentences per event, each containing two event elements. Thus, all possible pairings of event elements were presented. Some events contained an agentic element that was placed as the grammatical subject in an active sentence. Other events did not contain an agentic element, and thus only passive sentences were used for these events. We expected to find a stronger stochastic dependency of the retrieval of event elements for events with an agentic element than for events without an agentic element (Hypothesis 1). The experiment's design, hypothesis, and analysis plan were preregistered at https://osf.io/kts8p.

Methods

Participants

Participants were recruited from the Web (using various channels such as Social Media, mailing lists, blogs, and the online research platform SurveyCircle, 2021). They could join a lottery for winning vouchers of a total value of 80€ or receive course credit. An a priori power analysis with simulated data based on data by Schreiner et al. (2022) for detecting a small to medium difference between conditions (difference in event-specific trait variances of 0.75 according to the statistical procedure [see below], cf. Glas et al., 2000; Wang et al., 2002, assumed baseline event-specific trait variance of 1) with 80% power using one-tailed testing yielded a desired sample size of 40 participants. Due to the potential necessity of some data exclusion we increased the desired sample size by 20%, thus planning for a sample size of 48 participants. Because we needed to exclude more participants than anticipated we collected data from an additional 15 participants. Thus, we collected data of 63 participants. All participants provided online informed consent for their participation and publication of their data. We excluded 20 participants from the analyses because they did not pass both attention checks. Another three participants were excluded because they conducted the study on a smartphone on which a correct display of the experiment content could not be guaranteed. An additional participant was excluded because their browsing behavior suggested they interrupted the experiment frequently and for a longer duration. Thus, the final sample consisted of 39 participants (30 [77%] female, 32 [82%] students) with an average age of 29.6 years (*SD* = 9.8, range = 20-64). All participants indicated speaking German as their first language or fluently.

Design

The experiment employed a one-factorial (agency condition: agency vs. non-agency) within-subjects design. In the agency condition one event element served as the agent and was placed as the grammatical subject in active sentences. In the non-agency condition there was no agent and only passive sentences were used.

Material

Stimuli consisted of 144 German nouns representing three different stimulus types (common objects): 48 means of transportation (e.g., *bicycle*), 48 tools (e.g., *hammer*), and 48 foods (fruits and vegetables, e.g., *apple*). In addition, 48 verbs (e.g., *grab*) were used. An additional 24 nouns, 8 of each type, and 8 verbs were used as primacy buffers for preventing primacy effects. Stimuli were partly taken from Schreiner et al. (2022). We used three types of objects to avoid confounding with animacy (cf. Schreiner et al., 2022), because we wanted to dissociate agency from animacy effects. Using the stimuli, we randomly created 48 events for each participant, each consisting of a means of transportation, a tool, a food, and a verb. Events were randomly assigned to the 2 experimental conditions, resulting in 24 events per condition and 4 primacy buffer events per condition, which were presented first.

Procedure

The experiment was conducted online and implemented using lab.js (Henninger et al., 2021). Data collection was managed by JATOS (Lange et al., 2015). The procedure was based on the separated encoding paradigm (Horner et al., 2015; Horner & Burgess, 2014). In the learning phase of the experiment, participants were presented a sentence containing two event elements and the verb associated with the event in each trial. Sentences referring to the same event were presented interleaved with sentences referring to other events. The presentation order was randomized with the constraint of a minimum of two other-event trials being presented between two same-event trials. The experimental conditions were randomly distributed across learning trials. In the agency condition active sentences were used if the sentence contained the agent and passive sentences were used if it did not (see Figure 1A for an example). Each stimulus type (i.e., means of transportation, tool, or food) served as the agent equally often across events. In the non-agency condition only passive sentences were used (see Figure 1B for an example). In passive sentences it was randomized which of the event elements appeared in the first sentence position. Thus, there were three learning trials for each event and all possible associations for each event were presented. Encoding episodes were thus coherent (using closed-loop structures, cf. Horner et al., 2015; Horner & Burgess, 2014). Event elements in the agency condition were defined in terms of whether they were the agent or one of the non-agents in an event, yielding the associations agent – non-agent$_1$, agent – non-agent$_2$, and non-agent$_1$ – non-agent$_2$. Event elements in the non-agency condition were defined in terms of their stimulus type, yielding the associations means of transportation – tool, means of transportation – food, and tool – food. Each trial consisted of a 0.5-s fixation cross, a 6-s sentence presentation, and a 1.5-s blank screen (see Figure 1C). Primacy buffers were presented at the beginning of the learning phase to prevent primacy effects and were not used for the later test phase. The learning phase included an attention check after 50% of learning trials (not counting primacy buffer trials). Participants were asked to click a button within 10 s. After the learning phase, participants conducted a filler task in which they had to solve randomly generated math problems for three minutes to avoid recency effects.

In the subsequent test phase participants performed an incidental cued recognition forced-choice task. In each test trial, participants were first presented a 0.5-s fixation cross, followed by a 3-s presentation of the cue word (one of the event elements shown in the learning phase), displayed in the screen center. After another 0.5-s fixation-cross, the cue word was again displayed in the screen center and and six response alternatives were displayed in a hexagonal array around it (see Figure 1D). Participants had to select the target response alternative that belonged to the same event as the cue word. All response alternatives were of the same stimulus type (e.g., all tools) and distractors were randomly drawn from other events. All associations were tested, but only in one direction to avoid testing effects. Thus, for example, in the non-agency condition we either used the cue-target pairs means of transportation – tool, tool – food, and food – means of transportation or the cue-target pairs means of transportation – food, food – tool, and tool – means of transportation for a given event, but not both. The direction tested was balanced across events within conditions. Thus, there were three test trials per event and each event element type served as a cue and target equally often across events. The test phase consisted of three blocks and one association per event was tested in each block. The order of test trials in each block was randomized. The test phase included another attention check after 50% of test trials. Participants were asked to select the top left response option.

Data analysis

All analyses were conducted in the R Programming Environment (R Core Team, 2021) and we used the R packages *papaja* (version 0.1.0.9997, Aust & Barth, 2020) and *tinylabels* (version 0.2.2, Barth, 2021) for reporting. We used the conventional significance level of $\alpha = .05$ for the analyses. For the exploratory analysis of memory performance we computed Bayes factors in favor of an effect. Thus, a Bayes factor > 1 is in favor of an effect, whereas a Bayes factor < 1 is in favor of the absence of an effect (see Jeffreys, 1961).

Exploratory analysis of memory performance. For an exploratory analysis of memory performance we used Bayesian generalized linear mixed models with a logit link function (Goldstein, 2011; Rouder & Lu, 2005). Test trial outcomes (i.e., whether a correct response was given by selecting the target or an incorrect response was given by selecting a distractor in the cued recognition test) served as a binary dependent variable. Thus, individual trial information, rather than aggregate information, was entered into the model

Figure 1

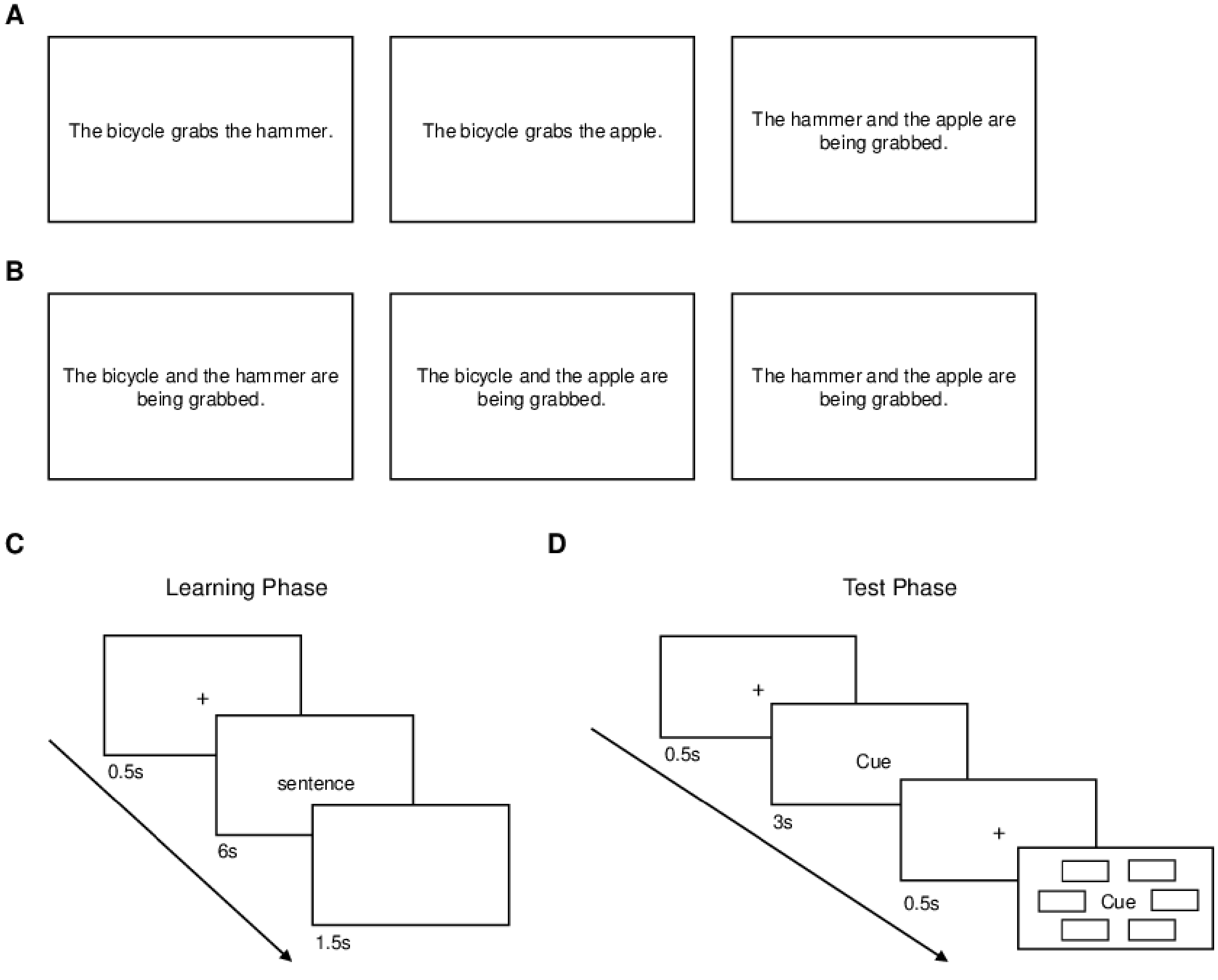
Experimental Procedure of Experiment 1

(A) Example Presentation of an Event in the Agency Condition

(B) Example Presentation of an Event in the Non-Agency Condition

(C) Schematic Depiction of a Learning Trial

(D) Schematic Depiction of a Test Trial



Note. Learning trials referring to the same event did not directly follow each other but were interleaved with at least two learning trials referring to other events.

(see Hoffman & Rovine, 2007). We investigated effects of agency condition, association[1], and the interaction. We also included random person intercepts to account for repeated measurement. Because, in the agency-condition, association refers to the agent or non-agent status of the cue and the target, whereas in the non-agency condition it refers to stimulus type (i.e., the factor association has different levels in the agency and non-agency condition), associations needed to be equated across agency conditions to jointly include them in the models. We equated corresponding factor levels, thus coercing the associations agent – non-agent$_1$ and means of transportation – tool, agent – non-agent$_2$ and means of transportation – food, and non-agent$_1$ – non-agent$_2$ and tool – food into a common factor level, respectively. To assess the influence of each factor, we fit several models with different predictors and compared them with a baseline model. To investigate the main effects, we compared a model including the respective predictor (condition or association) with a null model including only fixed and random person intercepts. To investigate the interaction, we compared the full model containing both main effects and the interaction with a model including both main effects but no interaction. We then computed Bayes factors in favor of an effect (BF_{10}) for each predictor.

Models were fit and Bayes factors were computed using the R package *brms* (version 2.16.4, Bürkner, 2017, 2018) using a standard normal prior for fixed effects and a half Student-t prior with three degrees of freedom (the default) for random effects. As a robustness check we also fit the models with less informative normal priors ($SD = 4$) and more informative normal priors ($SD = 0.25$) for the fixed effects and report the Bayes factors computed on the basis of these models in brackets behind the Bayes factors computed on the basis of the models with standard normal priors for the fixed effects. Models were fit with 4 Markov chains and 30,000 iterations per chain, the first 15,000 of

---

[1] Association refers to the element pair being tested. There were three associations tested in each condition. Associations do not distinguish between the direction of testing (e.g., the cue-target pairs tool – food and food – tool in the non-agency condition both test the association tool – food).

which were used as burnin iterations.

Dependency analysis. For modeling the stochastic dependency of the retrieval of event elements we used the approach by Schreiner et al. (2022) and Schreiner and Meiser (2022), which is based on item response theory (IRT, Lord, 1980; Lord & Novick, 1968). Thus, we fit a simplified three-parameter logistic IRT model (A. Birnbaum, 1968) to the data, with discrimination parameters fixed to 1, since events were randomly generated, and guessing parameters fixed to the stochastic guessing probability of $\frac{1}{6}$ given six response options in the cued recognition test:

$$P(u_{\text{ij}} = 1) = \frac{1}{6} + \frac{5}{6}\frac{e^{\theta_{\text{i}} - \beta_{\text{j}}}}{1 + e^{\theta_{\text{i}} - \beta_{\text{j}}}} \tag{1}$$

It models the probability of person i to give a correct response to item j given a latent person trait θ, representing memory performance in the current application, and item difficulty β. Based on this model we computed item residual correlations using the Q_3 statistic (Yen, 1984) with a bias correction (Yen, 1993) applied. The dependency measure D is then computed by contrasting the mean residual correlation between item pairs referring to the same event (*kk'*) with the mean residual correlation between item pairs referring to different events (*ll'*):

$$D = \frac{1}{K}\sum_{k > k'} Q_3^{\text{kk'}} - \frac{1}{L}\sum_{l > l'} Q_3^{\text{ll'}} \tag{2}$$

with K being the total number of item pairs referring to the same events and L being the total number of item pairs referring to different events. The model in (1) assumes local independence, meaning that the latent person trait accounts for all inter-item relationships (de Ayala, 2009; Lazarsfeld & Henry, 1968). Binding effects would violate this assumption and as a consequence residual correlations between item pairs referring to the same event would deviate from zero. They are contrasted with the mean residual correlation between item pairs referring to different events to control for baseline dependencies in empirical data and isolate the dependency that is specifically due to items being associated with a common event. This also makes D robust against model misspecification, since this would affect both the item residual correlations referring to the same event and item residual correlations referring to different events.

The approach requires parametric bootstrapping for obtaining p values (see Schreiner et al., 2022; Schreiner & Meiser, 2022), because the sampling distribution of Q_3, and thus also the one of D, is unknown (Chen & Thissen, 1997). To test whether dependency estimates differed from zero we repeatedly sampled from the model in (1), which assumes no dependency, using the empirically estimated item parameters. Person parameters were drawn from a normal distribution with a mean of zero and the empirically estimated latent trait variance. We then computed D for each sample and condition and used the resulting distributions for computing two-tailed p values and standard errors. To test whether dependency estimates differed between experimental conditions we repeatedly sampled from a bifactor IRT model (see Gibbons & Hedeker, 1992; Wainer & Wang, 2000), which extends the model in (1) by additional event-specific latent traits that exert their influence via their variance, thus inducing stochastic dependencies between items of the same event. Item parameters were empirically estimated by fitting a bifactor IRT model to the data. Person parameters were drawn from a multivariate normal distribution with means and covariances of zero and empirically estimated variances. Because events were randomly generated we set equality constraints on the event-specific trait variances within conditions when fitting the bifactor model. When sampling from the model, we set the event-specific trait variances of both conditions equal to the one of the condition with the smaller event-specific trait variance so the model assumes no difference in dependency between conditions. We then computed the difference of D between conditions for each sample and used the resulting distribution for computing a one-tailed p value and the standard error.

The R package *mirt* (version 1.35.1, Chalmers, 2012) and adapted functions from the package *sirt* (version 3.9-4, Robitzsch, 2020) were used for the dependency analysis.

The package *SimDesign* (version 2.8, Chalmers & Adkins, 2020) was used for conducting the parametric bootstraps. We used 1,000 bootstrap samples for each bootstrap (cf. Davison & Hinkley, 1997).

Results

Memory performance

On average, the proportion of correct responses was $M = 0.30$ ($SD = 0.46$) in the agency condition and $M = 0.27$ ($SD = 0.44$) in the non-agency condition. Figure 2 shows a raincloud plot (Allen et al., 2021) of the proportion of correct responses per participant. There was, depending on the choice of prior, weak evidence against to moderate evidence for a main effect of condition ($\mathrm{BF}_{10} = 1.87$ [0.48, 6.04]), but the 95% credible interval did not include zero, suggesting that memory performance was lower in the non-agency condition than in the agency condition ($\beta = -0.23$, $SE = 0.11$, 95% CI = [-0.44, -0.02], estimates are based on the full model). There was evidence against a main effect of association ($\mathrm{BF}_{10} = 0.005$ [< 0.001, 0.08]) and against an interaction of condition and association ($\mathrm{BF}_{10} = 0.03$ [0.002, 0.25]).

Dependency

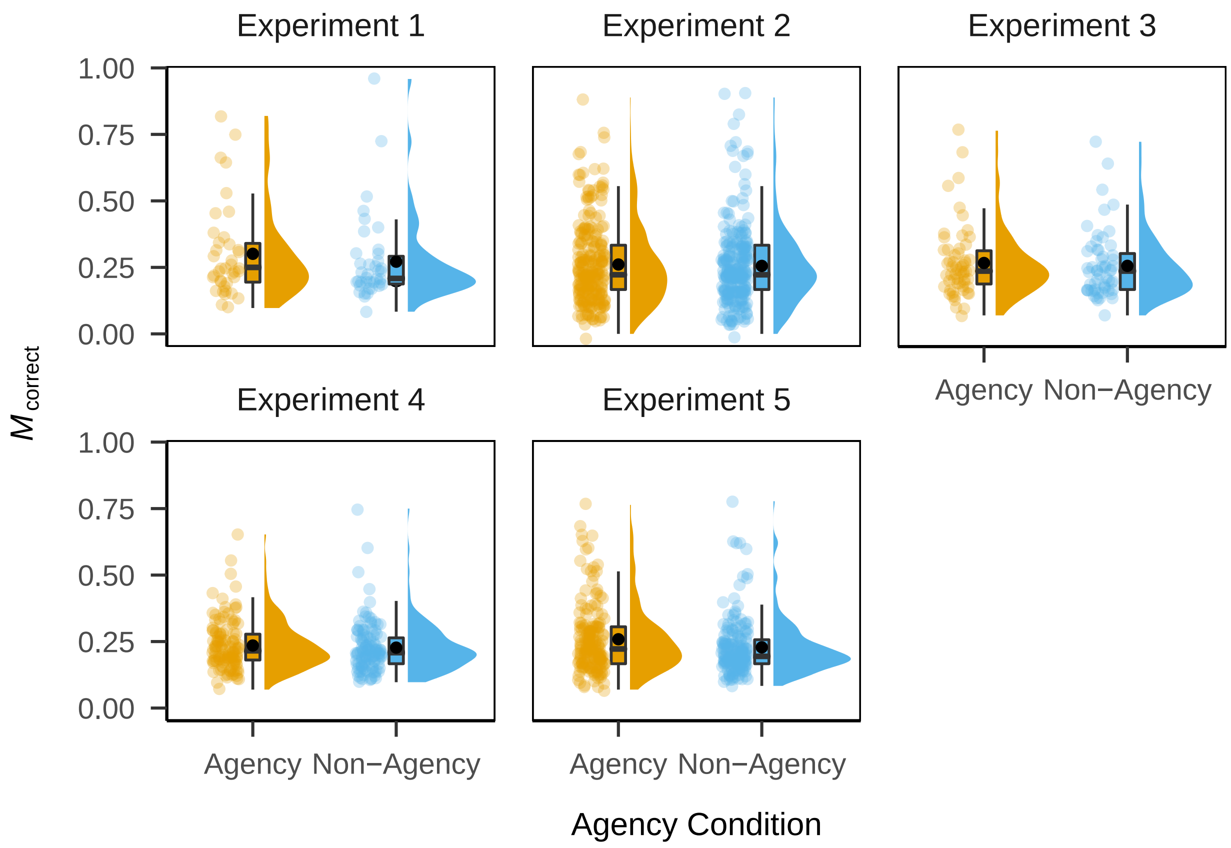
The dependency of the retrieval of event elements is shown in Figure 3. The dependency in both the agency condition ($D = 0.04$, $SE = 0.02$, $p = .05$) and non-agency condition ($D = 0.01$, $SE = 0.02$, $p = .61$) was non-significant. The dependency in the agency condition was not significantly larger than the one in the non-agency condition ($D_{\mathrm{diff}} = 0.03$, $SE = 0.03$, $p = .13$).

Discussion

In Experiment 1 we tested Hypothesis 1, which states that there is a stronger stochastic dependency of the retrieval of event elements for events with an agentic element than for events without an agentic element. There was no significant dependency in both the agency and non-agency condition and thus, the results are not particularly informative

Figure 2

Raincloud Plot Depticting the Proportion of Correct Responses per Participant by Agency Condition and Experiment

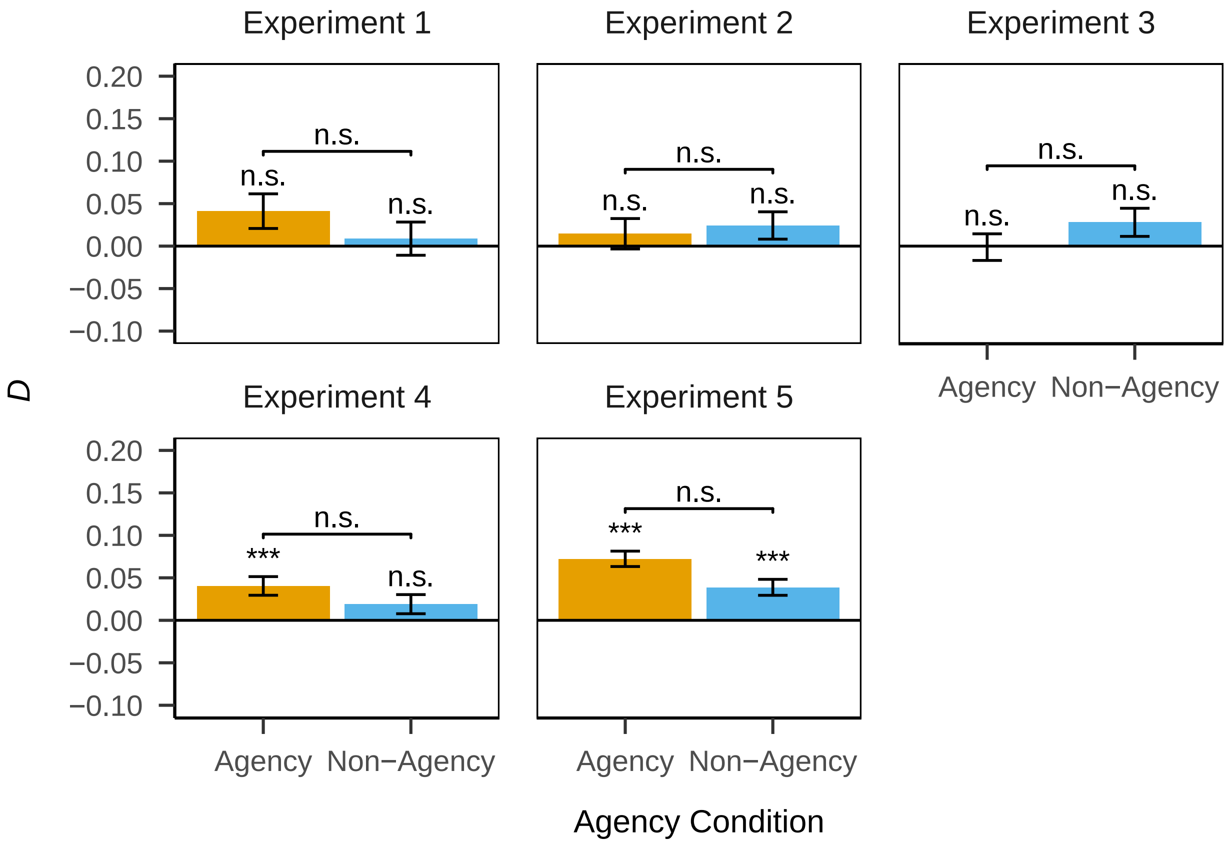


Notes. For Experiment 2 only the data of the closed-loop conditions are shown. Black dots depict the mean across participants.

regarding the hypothesis. However, the results descriptively pointed in the expected direction. There was a positive dependency in the agency condition that was larger than the dependency in the non-agency condition. Thus, the non-significant findings may be a Type II error or the effects may be rather small. In Experiment 2 we aimed for a larger sample size and thus for a higher power for detecting dependencies.

Figure 3

Dependency of the Retrieval of Event Elements by Agency Condition and Experiment



Notes. *** p < .001, n.s = non-significant. Error bars represent ± *SE*. For Experiment 2 only the data of the closed-loop conditions are shown.

Experiment 2

In Experiment 2 we again investigated whether there is a stronger stochastic dependency of the retrieval of event elements for events with an agentic than for events without an agentic element (Hypothesis 1). In addition, we investigated the binding structure of event elements by testing an integrated against a hierachical binding structure. To this end, we extended the experimental design to include non-coherent encoding episodes (open-loop structures, see Horner et al., 2015; Horner & Burgess, 2014) in addition to the coherent encoding episodes (closed-loop structures) that were used in Experiment 1.

While in closed-loop structures all possible pairwise associations between event elements are shown, we consistently excluded specific associations from presentation in the open-loop structures (cf. Schreiner et al., 2022). We report on this secondary research question in Appendix A and focus on the primary research question regarding agency effects on the binding of event elements in the main part of this article. The experiment's design, hypotheses, and analysis plan were preregistered at https://osf.io/kts8p.

Methods

Participants

Participants were again recruited from the Web, using the same channels as in Experiment 1. They could join a lottery for winning vouchers of a total value of 470€ or receive course credit. An a priori power analysis with simulated data for detecting the predicted results pattern with small to medium differences between conditions (difference in event-specific trait variances of 0.75 according to the statistical procedure, cf. Glas et al., 2000; Wang et al., 2002, assumed baseline event-specific trait variance of 1) with 80% power using one-tailed testing yielded a desired sample size of 240 participants. Due to the potential necessity of some data exclusion we increased the desired sample size by 20%, thus planning for a sample size of 288 participants. Because we needed to exclude more participants than anticipated we collected data from an additional 90 participants. Thus, we collected data of 378 participants. All participants provided online informed consent for their participation and publication of their data. We excluded three participants from the analyses because they indicated not speaking German fluently. Another 89 participants were excluded because they did not pass both attention checks. Another seven participants were excluded because they suggested their data should not be used for the study (e.g., due to distractions)[2]. Another 18 participants were excluded because they conducted the study

---

[2] We did not exclude participants who gave invalid exclusion reasons such as subjective bad memory performance. The exclusion reason of one of the participants suggested that the participant may not want their data to be used. We excluded the data of this participant from the data we made publicly available.

on a smartphone on which a correct display of the experiment content could not be guaranteed. Another six participants were excluded because they indicated having recently participated in a similar study (i.e., Experiment 1). An additional 13 participants were excluded because their browsing behavior suggested they interrupted the experiment frequently and for a longer duration. Thus, the final sample consisted of 242 participants (181 [75%] female, 1 [0.4%] non-binary, 198 [82%] students) with an average age of 27.1 years (*SD* = 9.6, range = 18-68).

Design

The experiment employed a 2 (agency condition: agency vs. non-agency) × 4 (loop condition: closed-loop and three open-loops) within-subjects design. The closed-loop conditions (CL) were identical to the conditions in Experiment 1. The open-loop conditions are described in Appendix A.

Material and procedure

Stimuli were identical to the ones of Experiment 1. Events were randomly assigned to the eight experimental conditions, resulting in six events per condition and one primacy buffer event per condition. The experimental procedure was identical to the one of Experiment 1, except that open-loop conditions were included in addition to closed-loop conditions.

Data analysis

Data analysis was identical to the one of Experiment 1 with the following exceptions: For the exploratory analysis of memory performance loop condition was included as an additional predictor. Thus, there were three possible two-way interactions and one possible three-way interaction. To investigate the three-way interaction we compared the full model including all main effects and interactions with a model containing all main effects and two-way interactions but no three-way interaction. Similar to the handling of the factor association we coerced loop conditions into common factor levels to jointly include them in the models for the analysis of memory performance (see

Appendix A for further information). To further investigate interactions we conducted post-hoc pairwise comparisons using the package *emmeans* (version 1.7.2, Lenth, 2022). We considered a difference to be substantial if the 95% credible interval (highest posterior density interval) does not include zero.

Results

Memory performance

On average, the proportion of correct responses was $M = 0.24$ ($SD = 0.42$) in the agency condition and $M = 0.22$ ($SD = 0.42$) in the non-agency condition. The proportion of correct responses by agency condition, loop condition, and association is shown in Figure B.1 in the Appendix. For loop conditions CL, Figure 2 additionally shows the proportion of correct responses per participant. There was, depending on the choice of prior, moderate evidence against to moderate evidence for a main effect of agency condition ($\mathrm{BF}_{10} = 1.12$ [0.27, 4.27]), but the 95% credible interval included zero, suggesting that memory performance did not substantially differ between the agency and non-agency condition ($\beta = -0.04$, $SE = 0.08$, 95% CI = [-0.20, 0.12], estimates are based on the full model). There was evidence for a main effect of loop condition ($\mathrm{BF}_{10} > 1{,}000$ [$> 1{,}000$, $> 1{,}000$]), but evidence against a main effect of association ($\mathrm{BF}_{10} = 0.006$ [< 0.001, 0.10]). There was moderate evidence against to evidence for a two-way interaction of agency condition and association ($\mathrm{BF}_{10} = 2.50$ [0.16, 21.28]). Post-hoc pairwise comparisons (see Table B.1 in the Appendix) revealed that, in the agency-condition, memory performance was higher for associations involving the agent than for the association not involving the agent, whereas memory performance did not differ between the two associations involving the agent. In the non-agency condition memory performance did not differ between associations. There was also evidence for a two-way interaction of loop condition and association ($\mathrm{BF}_{10} > 1{,}000$ [$> 1{,}000$, $> 1{,}000$]) that qualified the main effect of loop condition. Post-hoc pairwise comparisons (see Table B.2 in the Appendix) revealed that memory performance was lower for associations that were not presented during the learning phase (i.e., to-be-inferred associations) than for associations that were presented during the learning phase. There was evidence against a two-way interaction of agency condition and loop condition ($\mathrm{BF}_{10} = 0.002$ [< 0.001, 0.10]) and against a three-way interaction ($\mathrm{BF}_{10} = 0.002$ [< 0.001, 0.09]).

Dependency

The dependency of the retrieval of event elements in loop conditions CL is shown in Figure 3. There was no significant dependency of the retrieval of event elements in any of the conditions. In loop condition CL, dependencies were $D = 0.01$ ($SE = 0.02$, $p = .40$) in the agency condition and $D = 0.02$ ($SE = 0.02$, $p = .13$) in the non-agency condition. The dependency in loop condition CL in the agency condition was also not significantly larger than the one in the non-agency condition ($D_{\text{diff}} = -0.01$, $SE = 0.02$, $p = .70$). The results for the open-loop conditions are reported in Appendix A.

Discussion

In Experiment 2 we again tested Hypothesis 1, which states that there is a stronger stochastic dependency of the retrieval of event elements for events with an agentic element than for events without an agentic element. Since we did not find a significant dependency in any of the experimental conditions, the results are uninformative regarding the hypothesis. It is noteworthy that memory performance was even lower than in Experiment 1 and memory performance in both Experiment 1 and 2 was lower than, for example, memory performance in the experiments by Schreiner et al. (2022). Since low memory performance is associated with lower power for detecting dependencies and differences in dependencies (Schreiner & Meiser, 2022), in Experiment 3 we made some changes to the experimental design and procedure intended to improve memory performance.

Experiment 3

In Experiment 3 we again investigated whether there is a stronger stochastic dependency of the retrieval of event elements for events with an agentic than for events without an agentic element (Hypothesis 1). We introduced a number of changes to the design and procedure of Experiment 1 (and the closed-loop conditions of Experiment 2) intended to improve memory performance to achieve higher power for detecting binding effects (cf. Schreiner & Meiser, 2022). We changed the experimental design from a within- to a between-subjects design while keeping the number of events per condition identical to Experiment 1 (except for the primacy buffers). Thus, in total, participants saw only half the number of events as in Experiment 1 and consequently memory load was reduced. We further increased the presentation duration of the sentences. We also changed the stimuli from objects to animals, because using animate stimuli may lead to larger binding effects since they have an inherent agentic potential. This may also increase the plausibility of the described scenes and reduce processing costs due to semantic mismatches when combining the inherently non-agentic (or only weakly agentic) objects with an action verb (see Lowder & Gordon, 2015). The experiment's design, hypothesis, and analysis plan were preregistered at https://osf.io/vhmt4.

Methods

Participants

Participants were recruited via Prolific (https://www.prolific.co/) and received a compensation of £3.75. They were prescreened to be native German speakers and to not conduct the study on a smartphone. An a priori power analysis with simulated data for detecting a medium difference between conditions[3] (difference in event-specific trait variances of 1 according to the statistical procedure, cf. Glas et al., 2000; Wang et al., 2002, assumed baseline event-specific trait variance of 1) with 80% power using one-tailed testing yielded a desired sample size of 100 participants (50 participants per between-subjects condition). Due to the potential necessity of some data exclusion we increased the desired sample size by 20% and collected data from 120 participants. All participants provided

---

[3] We planned the sample size based on a larger effect in Experiment 3 than in the preceding experiments because we expected that the changes to the experimental design and procedure would lead to larger binding effects and we wanted to first evaluate the impact of the changes using a smaller sample so that we could later follow-up with a larger sample.

online informed consent for their participation and publication of their data. The data of one participant was not transmitted due to a technical error. Two participants were excluded because they suggested their data should not be used for the study (e.g., due to tiredness). Thus, the final sample consisted of 59 participants in the agency condition (31 [53%] female, 2 [3%] non-binary, 21 [36%] students), with an average age of 34.9 years (*SD* = 13.2, range = 18-69), and 58 participants in the non-agency condition (25 [43%] female, 27 [47%] students), with an average age of 33.3 years (*SD* = 11.3, range = 19-62).

Design

The experiment employed a one-factorial (agency condition: agency vs. non-agency) between-subjects design. Participants were randomly assigned to the experimental conditions. Conditions were identical to the ones in Experiment 1 and to the closed-loop conditions in Experiment 2.

Material and procedure

Stimuli consisted of 72 German nouns representing three different animal types: 24 mammals (e.g., *dog*, partly taken from Schreiner et al., 2022), 24 birds (e.g., *eagle*), and 24 insects[4] (e.g., *ant*). In addition, 24 verbs (a subset of the ones of Experiment1) were used. An additional nine nouns (three of each type) and three verbs were used as primacy buffers. Using three types of animals instead of objects still avoids confounding with animacy, assuming that animacy is constant across the different animal types. Using the stimuli, we randomly created 24 events and 3 primacy buffer events for each participant, each consisting of a mammal, a bird, an insect, and a verb.

The procedure was identical to the one of Experiment 1, except that we increased the duration of the sentence presentation during the learning phase to 8 s. Thus, a learning trial consisted of a 0.5-s fixation cross, an 8-s sentence presentation, and a 1.5-s blank screen. In addition, we substituted the attention checks with a procedure intended to keep

---

[4] The insects stimuli contained some animals that are not actually insects but are commonly perceived as such (e.g., *spider*). They belong to the more general category of arthropods.

participants engaged during the learning phase — they were asked to click on a continue button after the primacy buffer trials and after 25%, 50%, and 75% of learning trials[5].

Data analysis

Data analysis was identical to the one conducted in Experiment 1, with the following exceptions: For the exploratory analysis of memory performance, agency condition was used as a between- instead of a within-subjects factor in the Bayesian generalized linear mixed models. Consistent with our approach in Experiments 1 and 2 we coerced the associations agent – non-agent$_1$ and mammal – bird, agent – non-agent$_2$ and mammal – insect, and non-agent$_1$ – non-agent$_2$ and bird – insect into a common factor level, respectively. For the dependency analysis, we fit separate IRT models to the data of each agency condition. Fitting the bifactor IRT models yielded some extreme estimates for item parameters that, when being used as input for the parametric bootstrap for testing differences between the experimental conditions, caused item responses in the simulated data to have no variance. This prevented the estimation of item parameters for these items (cf. Snodgrass & Corwin, 1988) and consequently the computation of the dependency measure in the bootstrap. Instead of adjusting these extreme item parameters (two parameters [3%] in the model for the agency condition and seven parameters [10%] for the model in the non-agency condition) by a fixed constant (cf. Snodgrass & Corwin, 1988), we used a model-based approach to substitute them with random values drawn from the empirical distribution of the remaining parameters, using the *remp* function from the package *fishmethods* (version 1.11-3, Nelson, 2022).

---

[5] The number of participants taking more than 10 s to click on any of the continue buttons was quite low (6 in Experiment 3, 3 in Experiment 4, and 12 in Experiment 5 [one of which was excluded due to a very long response time]). In addition, we checked for conspicuous response patterns during the test phase.

Results

Memory performance

On average, the proportion of correct responses was $M = 0.27$ ($SD = 0.44$) in the agency condition and $M = 0.26$ ($SD = 0.44$) in the non-agency condition. The proportion of correct responses per participant is shown in Figure 2. There was evidence against main effects of condition ($\mathrm{BF}_{10} = 0.14$ [0.03, 0.47]) and association ($\mathrm{BF}_{10} = 0.04$ [0.003, 0.62]) and against an interaction of condition and association ($\mathrm{BF}_{10} = 0.02$ [0.001, 0.24]).

Dependency

The dependency of the retrieval of event elements is shown in Figure 3. The dependency in both the agency condition ($D = 0.00$, $SE = 0.02$, $p = .88$) and non-agency condition ($D = 0.03$, $SE = 0.02$, $p = .10$) was non-significant. The dependency in the agency condition was not significantly larger than the one in the non-agency condition ($D_{\mathrm{diff}} =$ -0.03, $SE = 0.03$, $p = .80$).

Discussion

In Experiment 3 we again tested Hypothesis 1, which states that there is a stronger stochastic dependency of the retrieval of event elements for events with an agentic element than for events without an agentic element. However, there was again no significant dependency in both the agency and non-agency condition and thus, the results are not informative regarding the hypothesis. In addition, the changes made to the experimental procedure and design in Experiment 3 did not lead to an increase in memory performance compared to Experiments 1 and 2. It is possible that the more diverse sample counteracted effects of the changes of the experimental design and procedure. Compared to Experiments 1 and 2 the sample in Experiment 3 was older and comprised more males and fewer students.

Interim Discussion

In Experiments 1-3 we did not find significant dependencies in any of the experimental conditions. Thus, it may be the case that the presentation of sentences in the context of the separated encoding paradigm hinders the formation of coherent memory representations and instead representations are composed of independent pairwise bindings. Previous studies found significant dependencies using the separated encoding paradigm (Horner et al., 2015; Horner & Burgess, 2014), at least if encoding episodes were coherent (Bisby et al., 2018; Horner et al., 2015; Horner & Burgess, 2014; Joensen et al., 2020; Schreiner et al., 2022). However, these studies used individual words or images as stimuli. Since stimuli were embedded in sentences in our experiments, the scene that participants were instructed to imagine was more prestructured and the described scenes were in parts rather nonsensical (likely more so in Experiments 1 and 2 than in Experiment 3). In addition, in our experiments all event elements belonged to the same superordinate category (objects or animals), whereas in previous experiments they belonged to different categories (e.g., animals, objects, and locations). Thus, event elements in our experiments were more semantically related. Presenting individual stimuli allows participants to freely associate them and may also reduce the prevalence of nonsensical scenes, since participants may try to construct scenes that make some sense to them, thus facilitating the formation of coherent memory representations. This may be further facilitated by using more semantically distinct event elements. Given the more prestructured presentation format in our experiments, with events being presented as sentences, participants may have relied on independent pairwise representations, perhaps with different exemplars of the same stimuli (e.g., two different types of dogs). Embedding event elements in sentences may thus not work well in combination with the separated encoding paradigm. Since our agency manipulation relies on this presentation format, the separated encoding paradigm may not be suitable to investigate effects of agency on the binding of event elements in the current research. In Experiment 4 we switched to the simultaneous encoding paradigm (Horner &

Burgess, 2013), in which all event elements are presented simultaneously. This paradigm tends to yield more robust binding effects (see James et al., 2020).

Experiment 4

In Experiment 4 we again investigated whether there is a stronger stochastic dependency of the retrieval of event elements for events with an agentic than for events without an agentic element (Hypothesis 1), but changed the experimental paradigm to the simultaneous encoding paradigm (Horner & Burgess, 2013). In the simultaneous encoding paradigm all event elements are presented simultaneously, in a single learning trial, instead of being presented sequentially pairwise across different learning trials, as is the case in the separated encoding paradigm (Horner et al., 2015; Horner & Burgess, 2014). Given that the presentation of event elements embedded in sentences describing scenes seemed to hinder the formation of coherent memory representations in the separated encoding paradigm, this problem may not occur in the simultaneous encoding paradigm, in which it is no longer necessary to build coherent memory representations across temporarily divided encoding episodes. The experiment's design, hypothesis, and analysis plan were preregistered at https://osf.io/q5tme.

Methods

Participants

Participants were recruited via Prolific (https://www.prolific.co/) and received a compensation of £2.50. They were prescreened to be native German speakers, to not conduct the study on a smartphone, and to not have participated in Experiment 3. An a priori power analysis with simulated data for detecting a medium difference between conditions (difference in event-specific trait variances of 1 according to the statistical procedure, cf. Glas et al., 2000; Wang et al., 2002, assumed baseline event-specific trait variance of 2)[6] with 80% power using one-tailed testing yielded a desired sample size of 200 participants (100 participants per between-subjects condition). Due to the potential necessity of some data exclusion we increased the desired sample size by 20% and collected data from 241 participants[7]. All participants provided online informed consent for their participation and publication of their data. Two participants were excluded because they suggested their data should not be used for the study (e.g., due to tiredness). Thus, the final sample consisted of 122 participants in the agency condition (59 [48%] female, 1 [1%] non-binary, 43 [35%] students), with an average age of 31.3 years (*SD* = 10.7, range = 18-62), and 117 participants in the non-agency condition (53 [45%] female, 43 [37%] students), with an average age of 32.0 years (*SD* = 11.3, range = 18-72).

Design

The experiment again employed a one-factorial (agency condition: agency vs. non-agency) between-subjects design. Conditions were identical to the ones in Experiments 1 and 3 and to the closed-loop conditions in Experiment 2.

Material and procedure

The material and procedure were identical to the ones of Experiment 3, except that we switched to the simultaneous encoding paradigm (Horner & Burgess, 2013). In the simultaneous encoding paradigm all event elements are presented in a single learning trial. Thus, participants were presented only one sentence per event that contained three event elements. In the agency condition there were thus one agent (grammatical subject) and

---

[6] We increased the assumed baseline event-specific trait variance from 1 in Experiments 1-3 to 2 in Experiment 4 because of the change to simultaneous event element presentation, for which higher dependencies may be expected. Previous empirical data based on the simultaneous encoding paradigm yielded higher event-specific trait variances (James et al., 2020; Ngo et al., 2019; see also Schreiner et al., 2022). A higher baseline event-specific trait variance is associated with reduced power for detecting differences between conditions (see Schreiner & Meiser, 2022) and thus increases the demand in sample size.

[7] Due to a participant error on Prolific the data of one additional participant was collected.

two patients (grammatical objects) in an active sentence (e.g., *The dog grabs the eagle and the ant.*), whereas in the non-agency condition there were three non-agentic elements in a passive sentence (e.g., *The dog, the eagle, and the ant are being grabbed.*). The sentence positions of the non-agentic elements were randomized.

Data analysis

Data analysis was identical to the one conducted in Experiment 3. Fitting the bifactor IRT models again yielded some extreme estimates for item parameters that caused item responses in the simulated data used for the parametric bootstrap for testing differences between the experimental conditions to have no variance. As in Experiment 3, we substituted these parameters (four parameters [6%] in the model for the agency and non-agency condition, respectively) with values randomly drawn from the empirical distribution of the remaining parameters.

Results

Memory performance

On average, the proportion of correct responses was $M = 0.24$ ($SD = 0.42$) in the agency condition and $M = 0.23$ ($SD = 0.42$) in the non-agency condition. The proportion of correct responses per participant is shown in Figure 2. There was evidence against a main effect of condition ($\mathrm{BF}_{10} = 0.08$ [0.02, 0.31]), evidence against to weak evidence for a main effect of association (depending on the choice of prior, $\mathrm{BF}_{10} = 0.16$ [0.01, 2.35]), and evidence against an interaction of condition and association ($\mathrm{BF}_{10} = 0.009$ [< 0.001, 0.12]).

Dependency

The dependency of the retrieval of event elements is shown in Figure 3. There was a significant positive dependency in the agency condition ($D = 0.04$, $SE = 0.01$, $p < .001$). The dependency in the non-agency condition was non-significant ($D = 0.02$, $SE = 0.01$, $p = .12$). Testing for a difference in dependency between conditions, the dependency in the agency condition was not significantly larger than the one in the non-agency condition

($D_{\text{diff}} = 0.02$, $SE = 0.03$, $p = .22$).

Discussion

In Experiment 4 we again tested Hypothesis 1, which states that there is a stronger stochastic dependency of the retrieval of event elements for events with an agentic element than for events without an agentic element. Contrary to Experiments 1-3, all event elements were presented simultaneously. The pattern of results, with a significant positive dependency in the agency condition and no significant dependency in the non-agency condition, is in favor of the hypothesis and suggests that the presence of an agentic element in an event facilitates the binding of event elements, leading to more coherent memory representations. However, the difference in dependency between conditions did not reach significance. Thus, Hypothesis 1 was only partially supported. It may be the case that the difference in dependency between conditions was smaller than anticipated and the study did thus not have enough power for reliably detecting the difference. In Experiment 5 we aimed to replicate the pattern of results and to detect also smaller differences between conditions by increasing the sample size.

Experiment 5

Experiment 5 was a replication of Experiment 4 and we thus again investigated whether there is a stronger stochastic dependency of the retrieval of event elements for events with an agentic than for events without an agentic element (Hypothesis 1). In Experiment 5 we aimed at a higher power for detecting smaller differences between the experimental conditions. We also slightly increased the duration of the learning trials to improve memory performance. The experiment's design, hypothesis, and analysis plan were preregistered at https://osf.io/g59uh.

Methods

Participants

Participants were recruited via Prolific (https://www.prolific.co/) and received a compensation of £2.63. They were prescreened to be native German speakers, to not conduct the study on a smartphone, and to not have participated in Experiments 3 and 4. An a priori power analysis (the same as in Experiment 4) for detecting a small to medium difference between conditions (difference in event-specific trait variances of 0.75 according to the statistical procedure, cf. Glas et al., 2000; Wang et al., 2002, assumed baseline event-specific trait variance of 2) with 80% power using one-tailed testing yielded a desired sample size of 300 participants (150 participants per between-subjects condition). Due to the potential necessity of some data exclusion we increased the desired sample size by 20% and collected data from 360 participants. All participants provided online informed consent for their participation and publication of their data. The data of one participant was not transmitted due to a technical error. One participant was excluded because they processed less than five math problems during the filler task. Another six participants were excluded because they suggested their data should not be used for the study (e.g., due to not properly understanding the instructions). An additional participant was excluded because their data suggested that they interrupted the study for a long duration of about nine minutes during the learning phase. Thus, the final sample consisted of 180 participants in the agency condition (98 [54%] female, 2 [1%] non-binary, 82 [46%] students), with an average age of 29.7 years (*SD* = 10.2, range = 18-70), and 171 participants in the non-agency condition (80 [47%] female, 3 [2%] non-binary, 75 [44%] students), with an average age of 31.2 years (*SD* = 10.3, range = 18-69).

Design, material, procedure, and data anlysis

The experimental design, the stimuli, and the data analysis were identical to the ones of Experiment 4. The experimental procedure was also identical to the one of Experiment 4 except that we increased the presentation duration of the sentences to 10 s and the duration of the blank screen to 2 s. Thus, each learning trial consisted of a 0.5-s fixation cross, a 10-s sentence presentation, and a 2-s blank screen.

Results

Memory performance

On average, the proportion of correct responses was $M = 0.26$ ($SD = 0.44$) in the agency condition and $M = 0.23$ ($SD = 0.42$) in the non-agency condition. The proportion of correct responses per participant is shown in Figure 2. There was, depending on the choice of prior, moderate evidence against to weak evidence for a main effect of condition ($\mathrm{BF}_{10} = 1.16$ [0.31, 3.70]), but the 95% credible interval did not include zero, suggesting that memory performance was lower in the non-agency condition than in the agency condition ($\beta = -0.18$, $SE = 0.08$, 95% CI = [-0.33, -0.02], estimates are based on the full model). There was evidence against a main effect of association ($\mathrm{BF}_{10} = 0.11$ [0.007, 1.57]) and against an interaction of condition and association ($\mathrm{BF}_{10} = 0.03$ [0.002, 0.39]).

Dependency

The dependency of the retrieval of event elements is shown in Figure 3. There was a significant positive dependency in both the agency condition ($D = 0.07$, $SE = 0.01$, $p < .001$) and the non-agency condition ($D = 0.04$, $SE = 0.01$, $p < .001$). Testing for a difference in dependency between conditions, the dependency in the agency condition was not significantly larger than the one in the non-agency condition ($D_{\mathrm{diff}} = 0.03$, $SE = 0.02$, $p = .15$).

Discussion

In Experiment 5 we largely replicated the findings of Experiment 4, except that the dependency in the non-agency condition reached significance in Experiment 5. This may be due to increased power given the larger sample size. Dependencies in both conditions were also descriptively larger than in Experiment 4, which may be due to the increased learning trial duration, which may have given participants more time to form bindings during encoding. While the difference in dependency between the agency and non-agency condition was descriptively larger than in Experiment 4, it did not reach significance, despite having a larger power for also detecting smaller effects in Experiment 5. Thus, Hypothesis 1, which states that there is a stronger stochastic dependency for events with an agentic than for events without an agentic element, was not supported. However, descriptively the results pointed in the expected direction, as was the case in Experiment 4. It may be the case that the relatively poor memory performance in the experiment makes it hard to detect significant differences in dependency between the conditions (cf. Schreiner & Meiser, 2022).

Aggregate Dependency Analysis with High Performers

To follow up on the supposition that low memory performance may have concealed the difference in dependency between the agency and non-agency condition in Experiments 4 and 5, we performed a post-hoc supplemental analysis in which we only included participants with above-median performance (i.e., above-median individual proportion of correct responses in the memory test in the respective condition) and aggregated the data of Experiments 4 and 5 to account for the reduced sample sizes. The resulting sample sizes were $N = 149$ in the agency condition and $N = 136$ in the non-agency condition. In favor of our supposition, this analysis yielded a significant difference in dependency between the agency and non-agency condition ($D_{\text{diff}} = 0.04$, $SE = 0.02$, $p = .03$), with a significant positive dependency in both the agency ($D = 0.09$, $SE = 0.01$, $p < .001$) and non-agency ($D = 0.05$, $SE = 0.01$, $p < .001$) condition[8]. It may thus have been the case that the

---

[8] When running this analysis separately for Experiments 4 and 5, the difference in dependency between the agency and non-agency condition was significant in Experiment 5 ($D_{\text{diff}} = 0.06$, $SE = 0.03$, $p = .03$, agency condition: $N = 88$, $D = 0.11$, $SE = 0.01$, $p < .001$, non-agency condition: $N = 82$, $D = 0.05$, $SE = 0.01$, $p < .001$), but not in Experiment 4 ($D_{\text{diff}} = 0.02$, $SE = 0.03$, $p = .31$, agency condition: $N = 61$, $D = 0.07$, $SE = 0.02$, $p < .001$, non-agency condition: $N = 54$, $D = 0.05$, $SE = 0.02$, $p = .002$), but note the reduced sample sizes.

rather poor memory performance in the experiments somewhat concealed the differences in dependency between the conditions or that there was a rather large group of low-performing participants (e.g., 18-22% of participants performed below chance level in Experiment 4 and 19-20% did so in Experiment 5) for which binding was mostly unsuccessful, which diluted the effect of agency.

General Discussion

In five experiments we investigated whether the binding of event elements in episodic memory is influenced by agency, using a linguistic agency manipulation. While the results of this research cannot give a definite answer to the research question, they nevertheless hint at a moderating role of agency. In addition, results strongly diverged between experiments in which event elements were presented sequentially pairwise or simultaneously, suggesting an effect of the experimental paradigm. An overview of the experimental setups and results is given in Table 1.

Table 1

Overview of Methods Aspects and Results for all Experiments

	Experiment				
	1	2	3	4	5
Encoding	separated	separated	separated	simultaneous	simultaneous
Design	within	within	between	between	between
Material	objects	objects	animals	animals	animals
Conditions	2	8	2	2	2
Events (total)$^+$	48	48	24	24	24
Events (per condition)$^+$	24	6	24	24	24
Learning trial duration	8 s	8 s	10 s	10 s	12.5 s
Data collection	Web	Web	Prolific	Prolific	Prolific
N	39	242	59/58	122/117	180/171
D in agency condition	non-sign.	non-sign.	non-sign.	sign.	sign.
D in non-agency condition	non-sign.	non-sign.	non-sign.	non-sign.	sign.
D_{diff}	non-sign.	non-sign.	non-sign.	non-sign.	non-sign.

Notes. $^+$not including primacy buffers. D = dependency of the retrieval of event elements, D_{diff} = difference in dependency of the retrieval of event elements between conditions, sign. = significant, non-sign. = non-significant. Learning trial duration encompasses the fixation cross, the sentence presentation, and the blank screen. Sample sizes divided by a slash (/) refer to the sample size in the agency condition and non-agency condition, respectively (for between-subjects designs). For Experiment 2 result information refers to the closed-loop conditions.

Free Association may Facilitate Binding Across Temporarily Divided Encoding Episodes

We only found significant dependencies of the retrieval of event elements when using the simultaneous encoding paradigm (Horner & Burgess, 2013), in which event elements are presented simultaneously (Experiments 4 and 5), but not when using the separated encoding paradigm (Horner et al., 2015; Horner & Burgess, 2014), in which event elements are presented sequentially pairwise (Experiments 1-3). This differs from previous findings, which showed significant dependencies of the retrieval of event elements also in the separated encoding paradigm, at least for coherent encoding episodes in which all possible pairwise associations are presented (Bisby et al., 2018; Horner et al., 2015; Horner & Burgess, 2014; Joensen et al., 2020; Schreiner et al., 2022). Some studies even found comparable dependencies between separated and simultaneous encoding conditions (Bisby et al., 2018; Horner & Burgess, 2014). However, in all of these studies event elements were presented as individual words or images, whereas we presented event elements embedded in sentences. In addition, our event elements were more semantically related, all of them being either objects or animals, than in the other studies in which event elements belonged to more distinct categories (e.g., animals, objects, and locations). In all of the studies participants were required to imagine the event elements as part of a scene and imagine them interacting in a meaningful manner. This may be easier if event elements are presented as individual words or images and are more semantically distinct because this allows participants to more freely associate them than if they are presented in a more guided fashion and are more semantically similar. When freely associating event elements, participants may also try to construct scenes that make sense to them, whereas the sentences presented in our experiments may have made less sense to the participants. Thus, the free association of event elements may facilitate the formation of coherent memory representations, whereas a more prestructured presentation of events may have caused participants to rely on independent pairwise representations. This adds to previous research suggesting that additional processes are required when binding event elements across temporarily divided encoding episodes compared to binding them within a single encoding episode (James et al., 2020). Besides written (rather than pictorial) and unidimensional (rather than multidimensional) presentation of event elements (James et al., 2020), the possibility to freely associate them may facilitate binding. Thus, the agency manipulation we employed, in which event elements are embedded in sentences, may not work well in combination with the separated encoding paradigm used in Experiments 1-3.

Agency may Facilitate the Binding of Event Elements

When using the simultaneous encoding paradigm we found significant dependencies of the retrieval of event elements. In Experiment 4 this was only the case in the agency condition. In Experiment 5 we found significant dependencies in both the agency and non-agency condition. While the difference in dependency between conditions did not reach significance in both experiments, the results descriptively pointed in the expected direction, with a higher dependency of the retrieval of event elements in the agency than in the non-agency condition. It may have been the case that the rather poor memory performance in the experiments concealed a potential effect of agency, since it is harder to find differences in dependency between conditions at lower levels of memory performance (cf. Schreiner & Meiser, 2022). In the aggregate dependency analysis of Experiments 4 and 5, in which we only included participants who performed above the median, the difference between the agency and non-agency condition was significant, suggesting that the non-significant difference in Experiments 4 and 5 may indeed be due to low memory performance. Thus, while the results do not allow for a definite answer to the question whether agency facilitates the binding of event elements, they hint at a facilitating effect of agency.

This extends previous findings on effects of agency in relation to memory and cognition. While agency has been found to influence visual attention (Huffman & Brockmole, 2020; Wen & Haggard, 2018), object vision (Haxby et al., 2020; Thorat et al., 2019), face perception (Walker & Keller, 2019), autobiographical memory (Woike et al., 1999; Woike & Polo, 2001), hippocampal responses to expectation violations in episodic memory (Jainta et al., 2022), and memory performance (Hon & Yeo, 2021; Hornstein & Mulligan, 2001), the current research suggests that agency may also play a role in the formation of memory representations in episodic memory by influencing binding processes. This may also provide a more proximate explanation for previously found facilitating effects of animacy (e.g., Bonin et al., 2015; Nairne et al., 2013, 2017) on the binding of event elements (Schreiner et al., 2022).

There are several potential explanations for effects of agency on the binding of event elements. Animacy effects regarding memory performance (e.g., Li et al., 2016; Nairne et al., 2013; VanArsdall et al., 2015) have been commonly explained by animacy being an important survival-related factor (Nairne et al., 2013, 2017) that influences a memory system that has evolved as a product of selective pressure on our ancestors (Nairne et al., 2007, 2008). Following this reasoning, agency may also be an important survival-related factor. For example, agentic individuals may be particularly dangerous opponents or, considering the role of agency in status perception (e.g., Carrier et al., 2014; Conway et al., 1996), agentic individuals may be associated with power and the ability to distribute rewards and punishment. This may explain why the presence of an agentic element in an event facilitates the formation of coherent memory representations, because such events may be particularly important to remember in a coherent manner. On a more perceptual level, attention may be biased towards the agentic element. A sense of agency over a stimulus has been shown to bias visual attention towards the stimulus which is under the participant's control (Huffman & Brockmole, 2020; Wen & Haggard, 2018). In addition, attention is biased towards dominant faces (Jones et al., 2010; Maner et al., 2008) and dominance is associated with agency (e.g., Abele & Wojciszke, 2007). Agents (and grammatical subjects, which the agentic element in our experiments always was) are also perceived as being more salient than patients (and grammatical objects, Kasof & Lee, 1993; Prat-Sala & Branigan, 2000). Finally, the sentences in our agency conditions provided a closed causal structure, in which both the agent and the patients of the described action were known. In contrast, in our non-agency conditions, only the patients of the described action were known, whereas the agent was unknown. It may be the case that the presence of an agentic element allows for the formation of closed causal structures, which facilitate the formation of coherent memory representations.

Ambiguous Evidence for Effects of Agency on Memory Performance

We found ambiguous evidence for effects of agency on memory performance. While there was evidence against an effect of agency in Experiments 3 and 4, there was some ambiguous evidence for an effect in Experiments 1, 2, and 5, as evaluated by the Bayes factor, that was sensitive to the choice of prior. However, the parameter estimates and 95% credible intervals in Experiments 1 and 5 suggested that memory performance was lower in the non-agency than in the agency condition. In addition, there was some evidence for an interaction between agency condition and association in Experiment 2 (but this was also sensitive to the choice of prior), suggesting that memory performance was higher for associations involving an agent than associations not involving an agent. Together these findings suggest that effects of agency on memory performance may have been present in some of the experiments but tended to be rather weak.

The findings from Experiments 1, 2 and 5 add to some initial evidence relating agency to memory performance, with evidence that a sense of agency over a stimulus improves memory performance for that stimulus (Hon & Yeo, 2021) and that self-performed tasks are remembered better than tasks performed by others (Hornstein & Mulligan, 2001). Adopting the view that agency effects may be due to survival-relatedness, similar to effects of animacy, and the finding that animacy advantages in memory performance were found reliably and across different test formats (e.g., Bonin et al., 2014, 2015; DeYoung & Serra, 2021; Leding, 2019, 2020; Li et al., 2016; Nairne et al., 2013; VanArsdall et al., 2013, 2015), one may expect a similar effect for agency. However, such an animacy effect has not been found using the cued recognition test used in the current research (Schreiner et al., 2022). In addition, Bonin et al. (2015) found that an imagery instruction, which we used in the current research, improves performance for inanimate but not for animate words. This may also apply to agentic and non-agentic elements.

Limitations

There are at least four limitations concerning the current research. First, all event elements in our experiments were either objects or animals, belonging to different subcategories of these classes (e.g., tools or mammals). We did this to control for potential confounding effects of animacy (cf. Schreiner et al., 2022). However, a drawback of using stimuli from the same superordinate category is that they are more semantically similar than when using stimuli from different categories (e.g., animals and locations). Thus, the stimuli in our experiments may have been harder to discriminate than stimuli used in previous studies (Bisby et al., 2018; Horner et al., 2015; Horner & Burgess, 2013, 2014; James et al., 2020; Joensen et al., 2020; Schreiner et al., 2022), which may have affected the results. In addition, not all the animal subcategories we used may be perceived as equal regarding animacy. Insects may be associated with lower animacy than birds or mammals (Connolly et al., 2012; Sha et al., 2015; Thorat et al., 2019). Thus, in Experiments 3-5 there may still have been some confounding with animacy, although it was certainly lower than if we had used more discrete categories such as animals, objects, and locations. Using animals from the same subcategory however, would have further increased the semantic similarity of the stimuli and further decreased their discriminability. In addition, the results of the memory performance analyses indicated no difference in memory performance for the different subcategories, which makes effects of the study material seem unlikely, at least regarding memory performance.

Second, memory performance in all experiments was rather low, to which a reduced stimulus discriminability may have contributed due to using stimuli from the same superordinate category. Lower memory performance is associated with lower power for detecting dependencies and differences in dependency between conditions (Schreiner & Meiser, 2022). Measures taken to increase memory performance, such as increasing the duration of the learning trials and reducing the total number of events each participant had to learn, proved unsuccessful. This may also explain why the differences in dependency between the agency and non-agency conditions in Experiments 4 and 5 did not reach significance.

Third, while there was no explicit agent in the non-agency condition (i.e., the agent was unknown), one may argue that participants may have *imagined* an agent and thus that there may have been an implicit agentic element in the non-agency condition. While this is a possibility, it seems inconsistent with the descriptive result patterns of Experiments 4 and 5 and the results of the aggregate dependency analysis, since, given a facilitating effect of agency on the binding of event elements, this should have boosted dependency in the non-agency condition. Rather, this line of reasoning suggests that effects of agency may actually be underestimated in the current experiments, if participants indeed imagined additional agentic elements in the non-agency condition.

Finally, all experiments were conducted online, with convenience Web samples for Experiments 1 and 2 and crowdsourced samples using Prolific for Experiments 3-5. While Web-based studies do not allow for the same degree of experimental control as do studies in the laboratory, several studies yielded comparable data quality for Web- and lab-based studies (Armitage & Eerola, 2020; Bartneck et al., 2015; Dandurand et al., 2008; de Leeuw & Motz, 2016; Hilbig, 2016) and participants' attention does not necessarily decrease during Web-based studies (Clifford & Jerit, 2014; Hauser & Schwarz, 2016). We also employed rigorous data quality checks in our experiments. In addition, replicable effects for research on the binding of event elements have been found in Web-based studies before (James et al., 2020; Schreiner et al., 2022). Thus, it is unlikely that the Web-based setting of our experiments invalidates our results. Conducting the experiments online also allowed us to gather more diverse samples than in typical psychological lab-based studies (see also

M. H. Birnbaum & Reips, 2005; Mason & Suri, 2012).

Directions for Future Research

Because the current research does not provide a definite answer to the question whether agency facilitates the binding of event elements in episodic memory, this question requires further investigation. To this end, future research should try to replicate our findings and to use different agency manipulations than the linguistic manipulation used in the current research. A different agency manipulation may also work in combination with the separated encoding paradigm (Horner et al., 2015; Horner & Burgess, 2014), allowing to investigate the structure in which event elements are bound together (see Appendix A and Schreiner et al., 2022). Of course, future research could also use an alternative procedure for testing the binding structure of event elements. Future research should aim at boosting memory performance, since higher memory performance compared to the current research would make it easier to detect effects (cf. Schreiner & Meiser, 2022). It may also be interesting to investigate boundary conditions for effects of agency. For example, do agency effects only occur for animate elements that at least have the potential to act autonomously, or can they also occur for elements for which this is not the case, such as inanimate objects, given that agency may also be attributed to inanimate objects or entities (e.g., Johnson & Barrett, 2003; Lowder & Gordon, 2015)? Even if effects occur for inanimate elements, effects may still interact with animacy, perhaps increasing proportionally with an animacy continuum, such as the one existing in the ventral temporal cortex regarding object vision (Connolly et al., 2012; Sha et al., 2015; Thorat et al., 2019). Finally, the possibility that free association facilitates the binding of event elements across temporarily divided encoding episodes should be directly tested in future research.

Conclusion

In five experiments we tested whether the presence of an agentic element in an event facilitates the binding of event elements in episodic memory. While the results do not allow for a definite answer to this research question, they hint at a facilitating effect of agency. In addition, agency effects may have been concealed to some degree due to low memory performance in the experiments. Such agency effects may provide a more proximate explanation for previously found effects of animacy on the binding of event elements (Schreiner et al., 2022). In addition, the results suggest that the possibility to freely associate event elements may facilitate binding across temporarily divided encoding episodes.

Acknowledgements

We thank Alicia Gernand, Timo Seitz, Eva Pöhlmann, Bastian Schnieders, Celine Kumpf, Vanessa Gottschall, and Joséphine Meyer for their assistance in creating and translating the material and testing the experiments.

Funding: This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - GRK 2277 “Statistical Modeling in Psychology”.

References

Abele, A. E., Uchronski, M., Suitner, C., & Wojciszke, B. (2008). Towards an operationalization of the fundamental dimensions of agency and communion: Trait content ratings in five countries considering valence and frequency of word occurrence. *European Journal of Social Psychology*, *38*(7), 1202–1217. https://doi.org/10.1002/ejsp.575

Abele, A. E., & Wojciszke, B. (2007). Agency and communion from the perspective of self versus others. *Journal of Personality and Social Psychology*, *93*(5), 751–763. https://doi.org/10.1037/0022-3514.93.5.751

Allen, M., Poggiali, D., Whitaker, K., Marshall, T. R., Langen, J. van, & Kievit, R. A. (2021). Raincloud plots: A multi-platform tool for robust data visualization. *Wellcome Open Research*, 4:63. https://doi.org/10.12688/wellcomeopenres.15191.2

Armitage, J., & Eerola, T. (2020). Reaction time data in music cognition: Comparison of pilot data from lab, crowdsourced, and convenience Web samples. *Frontiers in Psychology*, *10*, 2883. https://doi.org/10.3389/fpsyg.2019.02883

Arnold, N. R., Heck, D. W., Bröder, A., Meiser, T., & Boywitt, C. D. (2019). Testing hypotheses about binding in context memory with a hierarchical multinomial modeling approach: A preregistered study. *Experimental Psychology*, *66*(3), 239–251. https://doi.org/10.1027/1618-3169/a000442

Aust, F., & Barth, M. (2020). *papaja: Prepare reproducible APA journal articles with R Markdown. R package version 0.1.0.9997.* https://github.com/crsh/papaja

Bandura, A. (1989). Human agency in social cognitive theory. *American Psychologist*, *44*(9), 1175–1184. https://doi.org/10.1037/0003-066X.44.9.1175

Barth, M. (2021). *Tinylabels: Lightweight variable labels. R package version 0.2.2.* https://cran.r-project.org/package=tinylabels

Bartneck, C., Duenser, A., Moltchanova, E., & Zawieska, K. (2015). Comparing the similarity of responses received from studies in Amazon's Mechanical Turk to studies conducted online and with direct recruitment. *PLOS ONE*, *10*(4), e0121595. https://doi.org/10.1371/journal.pone.0121595

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores.* Addison-Wesley.

Birnbaum, M. H., & Reips, U.-D. (2005). Behavioral research and data collection via the internet. In *Handbook of human factors in web design* (pp. 471–491). Erlbaum.

Bisby, J. A., Horner, A. J., Bush, D., & Burgess, N. (2018). Negative emotional content disrupts the coherence of episodic memories. *Journal of Experimental Psychology: General*, *147*(2), 243–256. https://doi.org/10.1037/xge0000356

Bock, J. K., & Warren, R. K. (1985). Conceptual accessibility and syntactic structure in sentence formulation. *Cognition*, *21*(1), 47–67. https://doi.org/10.1016/0010-0277(85)90023-X

Bonin, P., Gelin, M., & Bugaiska, A. (2014). Animates are better remembered than inanimates: Further evidence from word and picture stimuli. *Memory & Cognition*, *42*(3), 370–382. https://doi.org/10.3758/s13421-013-0368-8

Bonin, P., Gelin, M., Laroche, B., Méot, A., & Bugaiska, A. (2015). The "how" of animacy effects in episodic memory. *Experimental Psychology*, *62*(6), 371–384. https://doi.org/10.1027/1618-3169/a000308

Boywitt, C. D., & Meiser, T. (2012a). Bound context features are integrated at encoding. *Quarterly Journal of Experimental Psychology*, *65*(8), 1484–1501. https://doi.org/10.1080/17470218.2012.656668

Boywitt, C. D., & Meiser, T. (2012b). The role of attention for context-context binding of intrinsic and extrinsic features. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(4), 1099–1107. https://doi.org/10.1037/a0026988

Bröder, A. (2009). Semantically clustered words are stored with integrated context: Validating a measurement model for source memory, storage, and retrieval in free recall.

Zeitschrift Für Psychologie / Journal of Psychology, *217*(3), 136–148. https://doi.org/10.1027/0044-3409.217.3.136

Brown, R., & Fish, D. (1983). The psychological causality implicit in language. *Cognition*, *14*(3), 237–273. https://doi.org/10.1016/0010-0277(83)90006-9

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*, 1–28. 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), 395–411.

Cai, D. J., Aharoni, D., Shuman, T., Shobe, J., Biane, J., Song, W., Wei, B., Veshkini, M., La-Vu, M., Lou, J., Flores, S. E., Kim, I., Sano, Y., Zhou, M., Baumgaertel, K., Lavi, A., Kamata, M., Tuszynski, M., Mayford, M., … Silva, A. J. (2016). A shared neural ensemble links distinct contextual memories encoded close in time. *Nature*, *534*(7605), 115–118. https://doi.org/10.1038/nature17955

Carrier, A., Louvet, E., Chauvin, B., & Rohmer, O. (2014). The primacy of agency over competence in status perception. *Social Psychology*, *45*(5), 347–356. https://doi.org/10.1027/1864-9335/a000176

Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, *48*(6), 1–29. https://doi.org/10.18637/jss.v048.i06

Chalmers, R. P., & Adkins, M. C. (2020). Writing effective and reliable Monte Carlo simulations with the SimDesign package. *The Quantitative Methods for Psychology*, *16*(4), 248–280. https://doi.org/10.20982/tqmp.16.4.p248

Chen, W.-H., & Thissen, D. (1997). Local dependence indexes for item pairs using item response theory. *Journal of Educational and Behavioral Statistics*, *22*(3), 265–289. https://doi.org/10.3102/10769986022003265

Clifford, S., & Jerit, J. (2014). Is there a cost to convenience? An experimental comparison of data quality in laboratory and online studies. *Journal of Experimental Political*

Science, *1*(2), 120–131. https://doi.org/10.1017/xps.2014.5

Cohen, N. J., & Eichenbaum, H. (1993). *Memory, amnesia, and the hippocampal system.* MIT Press.

Connolly, A. C., Guntupalli, J. S., Gors, J., Hanke, M., Halchenko, Y. O., Wu, Y.-C., Abdi, H., & Haxby, J. V. (2012). The representation of biological classes in the human brain. *The Journal of Neuroscience*, *32*(8), 2608–2618. https://doi.org/10.1523/JNEUROSCI.5547-11.2012

Conway, M., Pizzamiglio, M. T., & Mount, L. (1996). Status, communality, and agency: Implications for stereotypes of gender and other groups. *Journal of Personality and Social Psychology*, *71*(1), 25–38. https://doi.org/10.1037/0022-3514.71.1.25

Damasio, A. R. (1989). The brain binds entities and events by multiregional activation from convergence zones. *Neural Computation*, *1*(1), 123–132. https://doi.org/10.1162/neco.1989.1.1.123

Dandurand, F., Shultz, T. R., & Onishi, K. H. (2008). Comparing online and lab methods in a problem-solving experiment. *Behavior Research Methods*, *40*(2), 428–434. https://doi.org/doi.org/10.3758/brm.40.2.428

Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their application.* Cambridge University Press.

de Ayala, R. J. (2009). *The theory and practice of item response theory.* Guilford Press.

de Leeuw, J. R., & Motz, B. A. (2016). Psychophysics in a Web browser? Comparing response times collected with JavaScript and Psychophysics Toolbox in a visual search task. *Behavior Research Methods*, *48*(1), 1–12. https://doi.org/10.3758/s13428-015-0567-2

DeYoung, C. M., & Serra, M. J. (2021). Judgments of learning reflect the Animacy advantage for memory, but not beliefs about the effect. *Metacognition and Learning.* https://doi.org/10.1007/s11409-021-09264-w

Eichenbaum, H. (1999). The hippocampus and mechanisms of declarative memory.

Behavioural Brain Research, *103*(2), 123–133. https://doi.org/10.1016/S0166-4328(99)00044-3

Fausey, C. M., & Boroditsky, L. (2010). Subtle linguistic cues influence perceived blame and financial liability. *Psychonomic Bulletin & Review*, *17*(5), 644–650. https://doi.org/10.3758/PBR.17.5.644

Gibbons, R. D., & Hedeker, D. R. (1992). Full-information item bi-factor analysis. *Psychometrika*, *57*(3), 423–436. https://doi.org/10.1007/bf02295430

Glas, C. A. W., Wainer, H., & Bradlow, E. T. (2000). MML and EAP estimation in testlet-based adaptive testing. In W. J. van der Linden & C. A. W. Glas (Eds.), *Computerized adaptive testing: Theory and practice* (pp. 271–287). Kluwer.

Gleitman, L. R., January, D., Nappa, R., & Trueswell, J. C. (2007). On the give and take between event apprehension and utterance formulation. *Journal of Memory and Language*, *57*(4), 544–569. https://doi.org/10.1016/j.jml.2007.01.007

Goldstein, H. (2011). *Multilevel statistical models* (4th ed.). Wiley.

Halicki, K. T., Ingendahl, M., Mayer, M., John, M., Schreiner, M. R., & Wänke, M. (2021). From which direction does the empire strike (back)? *Frontiers in Psychology*, *12*, 625554. https://doi.org/10.3389/fpsyg.2021.625554

Hauser, D. J., & Schwarz, N. (2016). Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior Research Methods*, *48*(1), 400–407. https://doi.org/10.3758/s13428-015-0578-z

Haxby, J. V., Gobbini, M. I., & Nastase, S. A. (2020). Naturalistic stimuli reveal a dominant role for agentic action in visual representation. *NeuroImage*, *216*, 116561. https://doi.org/10.1016/j.neuroimage.2020.116561

Healy, M. J., & Caudell, T. P. (2019). Episodic memory: A hierarchy of spatiotemporal concepts. *Neural Networks*, *120*, 40–57. https://doi.org/10.1016/j.neunet.2019.09.021

Henley, N. M., Miller, M., & Beazley, J. A. (1995). Syntax, semantics, and sexual violence: Agency and the passive voice. *Journal of Language and Social Psychology*, *14*(1-2), 60–84. https://doi.org/10.1177/0261927X95141004

Henninger, F., Shevchenko, Y., Mertens, U. K., Kieslich, P. J., & Hilbig, B. E. (2021). lab.js: A free, open, online study builder. *Behavior Research Methods*. https://doi.org/10.3758/s13428-019-01283-5

Hilbig, B. E. (2016). Reaction time effects in lab- versus Web-based research: Experimental evidence. *Behavior Research Methods*, *48*(4), 1718–1724. https://doi.org/10.3758/s13428-015-0678-9

Hitlin, S., & Elder, G. H. (2007). Time, self, and the curiously abstract concept of agency. *Sociological Theory*, *25*(2), 170–191. https://doi.org/10.1111/j.1467-9558.2007.00303.x

Hoffman, L., & Rovine, M. J. (2007). Multilevel models for the experimental psychologist: Foundations and illustrative examples. *Behavior Research Methods*, *39*(1), 101–117. https://doi.org/10.3758/BF03192848

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The Theory of Event Coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences*, *24*(5), 849–878. https://doi.org/10.1017/s0140525x01000103

Hon, N., & Yeo, N. (2021). Having a sense of agency can improve memory. *Psychonomic Bulletin & Review*, *28*(3), 946–952. https://doi.org/10.3758/s13423-020-01849-x

Horner, A. J., Bisby, J. A., Bush, D., Lin, W.-J., & Burgess, N. (2015). Evidence for holistic episodic recollection via hippocampal pattern completion. *Nature Communications*, *6*(1), 7462. https://doi.org/10.1038/ncomms8462

Horner, A. J., & Burgess, N. (2013). The associative structure of memory for multi-element events. *Journal of Experimental Psychology: General*, *142*(4), 1370–1383. https://doi.org/10.1037/a0033626

Horner, A. J., & Burgess, N. (2014). Pattern completion in multielement event engrams. *Current Biology*, *24*(9), 988–992. https://doi.org/10.1016/j.cub.2014.03.012

Hornstein, S. L., & Mulligan, N. W. (2001). Memory of action events: The role of objects in memory of self- and other-performed tasks. *The American Journal of Psychology*, *114* (2), 199–217. https://doi.org/10.2307/1423515

Huffman, G., & Brockmole, J. R. (2020). Attentional selection is biased towards controllable stimuli. *Attention, Perception, & Psychophysics*, *82* (5), 2558–2569. https://doi.org/10.3758/s13414-020-02004-3

Jainta, B., Siestrup, S., El-Sourani, N., Trempler, I., Wurm, M. F., Werning, M., Cheng, S., & Schubotz, R. I. (2022). Seeing what i did (not): Cerebral and behavioral effects of agency and perspective on episodic memory re-activation. *Frontiers in Behavioral Neuroscience*, *15*, 793115. https://doi.org/10.3389/fnbeh.2021.793115

James, E., Ong, G., Henderson, L., & Horner, A. J. (2020). Make or break it: Boundary conditions for integrating multiple elements in episodic memory. *Royal Society Open Science*, *7* (9), 200431. https://doi.org/10.1098/rsos.200431

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Clarendon Press.

Joensen, B. H., Gaskell, M. G., & Horner, A. J. (2020). United we fall: All-or-none forgetting of complex episodic events. *Journal of Experimental Psychology: General*, *149* (2), 230–248. https://doi.org/10.1037/xge0000648

Johnson, A. H., & Barrett, J. (2003). The role of control in attributing intentional agency to inanimate objects. *Journal of Cognition and Culture*, *3* (3), 208–217. https://doi.org/10.1163/156853703322336634

Jones, B. C., DeBruine, L. M., Main, J. C., Little, A. C., Welling, L. L. M., Feinberg, D. R., & Tiddeman, B. P. (2010). Facial cues of dominance modulate the short-term gaze-cuing effect in human observers. *Proceedings of the Royal Society B: Biological Sciences*, *277* (1681), 617–624. https://doi.org/10.1098/rspb.2009.1575

Kako, E. (2006). Thematic role properties of subjects and objects. *Cognition*, *101* (1), 1–42. https://doi.org/10.1016/j.cognition.2005.08.002

Kasof, J., & Lee, J. Y. (1993). Implicit causality as implicit salience. *Journal of Personality and Social Psychology*, *65* (5), 877–891. https://doi.org/10.1037/0022-3514.65.5.877

Kassin, S., & Lowe, C. (1979). On the use of single sentence descriptions of behavior in attribution research. *Social Behavior and Personality*, *7*(1), 1–8. https://doi.org/10.2224/sbp.1979.7.1.1

Kazanas, S. A., Altarriba, J., & O'Brien, E. G. (2020). Paired-associate learning, animacy, and imageability effects in the survival advantage. *Memory & Cognition*, *48*(2), 244–255. https://doi.org/10.3758/s13421-019-01007-2

Koch, A., Imhoff, R., Dotsch, R., Unkelbach, C., & Alves, H. (2016). The ABC of stereotypes about groups: Agency/socioeconomic success, conservative-progressive beliefs, and communion. *Journal of Personality and Social Psychology*, *110*(5), 675–709. https://doi.org/10.1037/pspa0000046

Kumaran, D., & Ludwig, H. (2013). Transitivity performance, relational hierarchy knowledge and awareness: Results of an instructional framing manipulation. *Hippocampus*, *23*(12), 1259–1268. https://doi.org/10.1002/hipo.22163

Kumaran, D., & McClelland, J. L. (2012). Generalization through the recurrent interaction of episodic memories: A model of the hippocampal system. *Psychological Review*, *119*(3), 573–616. https://doi.org/10.1037/a0028681

Lange, K., Kühn, S., & Filevich, E. (2015). "Just another tool for online studies" (JATOS): An easy solution for setup and management of web servers supporting online studies. *PLOS ONE*, *10*(6), e0130834. https://doi.org/10.1371/journal.pone.0130834

Laurino, J., & Kaczer, L. (2019). Animacy as a memory enhancer during novel word learning: Evidence from orthographic and semantic memory tasks. *Memory*, *27*(6), 820–828. https://doi.org/10.1080/09658211.2019.1572195

Lazarsfeld, P. F., & Henry, N. W. (1968). *Latent structure analysis.* Houghton Mifflin.

Leding, J. K. (2019). Adaptive memory: Animacy, threat, and attention in free recall. *Memory & Cognition*, *47*(3), 383–394. https://doi.org/10.3758/s13421-018-0873-x

Leding, J. K. (2020). Animacy and threat in recognition memory. *Memory & Cognition*, *48*(5), 788–799. https://doi.org/10.3758/s13421-020-01017-5

Lenth, R. (2022). *emmeans: Estimated marginal means, aka least-squares means. R* *package version 1.7.2.* https://cran.r-project.org/package=emmeans

Li, P., Jia, X., Li, X., & Li, W. (2016). The effect of animacy on metamemory. *Memory & Cognition*, *44*(5), 696–705. https://doi.org/10.3758/s13421-016-0598-7

Lord, F. M. (1980). *Applications of item response theory to practical testing problems.* Erlbaum.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores.* Addison-Wesley.

Lowder, M. W., & Gordon, P. C. (2015). Natural forces as agents: Reconceptualizing the animate–inanimate distinction. *Cognition*, *136*, 85–90. https://doi.org/10.1016/j.cognition.2014.11.021

Madan, C. R. (2021). Exploring word memorability: How well do different word properties explain item free-recall probability? *Psychonomic Bulletin & Review*, *28*(2), 583–595. https://doi.org/10.3758/s13423-020-01820-w

Maner, J. K., DeWall, C. N., & Gailliot, M. T. (2008). Selective attention to signs of success: Social dominance and early stage interpersonal perception. *Personality and Social Psychology Bulletin*, *34*(4), 488–501. https://doi.org/10.1177/0146167207311910

Marr, D. (1971). Simple memory: A theory for archicortex. *Philosophical Transactions of the Royal Society of London. B: Biological Sciences*, *262*(841), 23–81. https://doi.org/10.1098/rstb.1971.0078

Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods*, *44*(1), 1–23. https://doi.org/10.3758/s13428-011-0124-6

McClelland, J. L., O'Reilly, R. C., & McNaughton, B. L. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*(3), 419–457. https://doi.org/10.1037/0033-295x.102.3.419

Meiser, T., & Bröder, A. (2002). Memory for multidimensional source information.

Journal of Experimental Psychology: Learning, Memory, and Cognition, *28*(1), 116–137. https://doi.org/10.1037//0278-7393.28.1.116

Moeller, B., Pfister, R., Kunde, W., & Frings, C. (2019). Selective binding of stimulus, response, and effect features. *Psychonomic Bulletin & Review*, *26*(5), 1627–1632. https://doi.org/10.3758/s13423-019-01646-1

Moll, M., & Miikkulainen, R. (1997). Convergence-zone episodic memory: Analysis and simulations. *Neural Networks*, *10*(6), 1017–1036. https://doi.org/10.1016/S0893-6080(97)00016-6

Morton, N. W., Schlichting, M. L., & Preston, A. R. (2020). Representations of common event structure in medial temporal lobe and frontoparietal cortex support efficient inference. *Proceedings of the National Academy of Sciences*, *117*(47), 29338–29345. https://doi.org/10.1073/pnas.1912338117

Nairne, J. S., Pandeirada, J. N. S., & Thompson, S. R. (2008). Adaptive memory: The comparative value of survival processing. *Psychological Science*, *19*(2), 176–180. https://doi.org/10.1111/j.1467-9280.2008.02064.x

Nairne, J. S., Thompson, S. R., & Pandeirada, J. N. S. (2007). Adaptive memory: Survival processing enhances retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(2), 263–273. https://doi.org/10.1037/0278-7393.33.2.263

Nairne, J. S., VanArsdall, J. E., & Cogdill, M. (2017). Remembering the living: Episodic memory is tuned to animacy. *Current Directions in Psychological Science*, *26*(1), 22–27. https://doi.org/10.1177/0963721416667711

Nairne, J. S., VanArsdall, J. E., Pandeirada, J. N. S., Cogdill, M., & LeBreton, J. M. (2013). Adaptive memory: The mnemonic value of animacy. *Psychological Science*, *24*(10), 2099–2105. https://doi.org/10.1177/0956797613480803

Nelson, G. A. (2022). *fishmethods: Fishery science methods and models. R package version 1.11-3.* https://cran.r-project.org/package=fishmethods

Ngo, C. T., Horner, A. J., Newcombe, N. S., & Olson, I. R. (2019). Development of holistic episodic recollection. *Psychological Science*, *30*(12), 1696–1706. https://doi.org/10.1177/0956797619879441

Popp, E. Y., & Serra, M. J. (2016). Adaptive memory: Animacy enhances free recall but impairs cued recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(2), 186–201. https://doi.org/10.1037/xlm0000174

Prat-Sala, M., & Branigan, H. P. (2000). Discourse constraints on syntactic processing in language production: A cross-linguistic study in English and Spanish. *Journal of Memory and Language*, *42*(2), 168–182. https://doi.org/10.1006/jmla.1999.2668

Prentice, J. L. (1967). Effects of cuing actor vs cuing object on word order in sentence production. *Psychonomic Science*, *8*(4), 163–164. https://doi.org/10.3758/BF03331600

Preston, A. R., Shrager, Y., Dudukovic, N. M., & Gabrieli, J. D. E. (2004). Hippocampal contribution to the novel use of relational information in declarative memory. *Hippocampus*, *14*(2), 148–152. https://doi.org/10.1002/hipo.20009

R Core Team. (2021). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing.

Rissman, L., Woodward, A., & Goldin-Meadow, S. (2019). Occluding the face diminishes the conceptual accessibility of an animate agent. *Language, Cognition and Neuroscience*, *34*(3), 273–288. https://doi.org/10.1080/23273798.2018.1525495

Robitzsch, A. (2020). *sirt: Supplementary item response theory models. R package version 3.9-4*. https://cran.r-project.org/package=sirt

Rouder, J. N., & Lu, J. (2005). An introduction to Bayesian hierarchical models with an application in the theory of signal detection. *Psychonomic Bulletin & Review*, *12*(4), 573–604. https://doi.org/10.3758/BF03196750

Schreiner, M. R., & Meiser, T. (2022). Measuring binding effects in event-based episodic representations. *Behavior Research Methods*. https://doi.org/10.3758/s13428-021-01769-1

Schreiner, M. R., Meiser, T., & Bröder, A. (2022). The binding structure of event elements in episodic memory and the role of animacy. *Quarterly Journal of Experimental Psychology*. https://doi.org/10.1177/17470218221096148

Sha, L., Haxby, J. V., Abdi, H., Guntupalli, J. S., Oosterhof, N. N., Halchenko, Y. O., & Connolly, A. C. (2015). The animacy continuum in the human ventral vision pathway. *Journal of Cognitive Neuroscience*, *27*(4), 665–678. https://doi.org/10.1162/jocn_a_00733

Shohamy, D., & Wagner, A. D. (2008). Integrating memories in the human brain: Hippocampal-midbrain encoding of overlapping events. *Neuron*, *60*(2), 378–389. https://doi.org/10.1016/j.neuron.2008.09.023

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, *117*(1), 34–50. https://doi.org/10.1037/0096-3445.117.1.34

Starns, J. J., & Hicks, J. L. (2005). Source dimensions are retrieved independently in multidimensional monitoring tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(6), 1213–1220. https://doi.org/10.1037/0278-7393.31.6.1213

Starns, J. J., & Hicks, J. L. (2008). Context attributes in memory are bound to item information, but not to one another. *Psychonomic Bulletin & Review*, *15*(2), 309–314. https://doi.org/10.3758/PBR.15.2.309

Suitner, C., & Maass, A. (2016). Spatial agency bias: Representing people in space. In J. M. Olson & M. P. Zanna (Eds.), *Advances in Experimental Social Psychology* (Vol. 53, pp. 245–301). Academic Press. https://doi.org/10.1016/bs.aesp.2015.09.004

SurveyCircle. (2021). *Research website SurveyCircle. Published 2016.* https://www.surveycircle.com.

Thorat, S., Proklova, D., & Peelen, M. V. (2019). The nature of the animacy organization in human ventral temporal cortex. *eLife*, *8*, e47142. https://doi.org/10.7554/eLife.47142

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory* (pp. 381–403). Academic Press.

Tulving, E. (1983). *Elements of episodic memory.* Oxford University Press.

VanArsdall, J. E., Nairne, J. S., Pandeirada, J. N. S., & Blunt, J. R. (2013). Adaptive memory: Animacy processing produces mnemonic advantages. *Experimental Psychology*, *60*(3), 172–178. https://doi.org/10.1027/1618-3169/a000186

VanArsdall, J. E., Nairne, J. S., Pandeirada, J. N. S., & Cogdill, M. (2015). Adaptive memory: Animacy effects persist in paired-associate learning. *Memory*, *23*(5), 657–663. https://doi.org/10.1080/09658211.2014.916304

Wainer, H., & Wang, X. (2000). Using a new statistical model for testlets to score TOEFL. *Journal of Educational Measurement*, *37*(3), 203–220. https://doi.org/10.1111/j.1745-3984.2000.tb01083.x

Walker, M., & Keller, M. (2019). Beyond attractiveness: A multimethod approach to study enhancement in self-recognition on the Big Two personality dimensions. *Journal of Personality and Social Psychology*, *117*(3), 483–499. https://doi.org/10.1037/pspa0000157

Wang, X., Bradlow, E. T., & Wainer, H. (2002). A general Bayesian model for testlets: Theory and applications. *Applied Psychological Measurement*, *26*(1), 109–128. https://doi.org/10.1177/0146621602026001007

Wen, W., & Haggard, P. (2018). Control changes the way we look at the world. *Journal of Cognitive Neuroscience*, *30*(4), 603–619. https://doi.org/10.1162/jocn_a_01226

Woike, B., Gershkovich, I., Piorkowski, R., & Polo, M. (1999). The role of motives in the content and structure of autobiographical memory. *Journal of Personality and Social Psychology*, *76*(4), 600–612. https://doi.org/10.1037/0022-3514.76.4.600

Woike, B., & Polo, M. (2001). Motive-related memories: Content, structure, and affect. *Journal of Personality*, *69*(3), 391–415. https://doi.org/10.1111/1467-6494.00150

Wojciszke, B., Abele, A. E., & Baryla, W. (2009). Two dimensions of interpersonal attitudes: Liking depends on communion, respect depends on agency. *European Journal of Social Psychology*, *39*(6), 973–990. https://doi.org/10.1002/ejsp.595

Yen, W. M. (1984). Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Applied Psychological Measurement*, *8*(2), 125–145. https://doi.org/10.1177/014662168400800201

Yen, W. M. (1993). Scaling performance assessments: Strategies for managing local item dependence. *Journal of Educational Measurement*, *30*(3), 187–213. https://doi.org/10.1111/j.1745-3984.1993.tb00423.x

Appendix A

Investigation of the Binding Structure of Event Elements in Experiment 2

Theoretical Background

In Experiment 2 we additionally investigated whether event elements are bound in an integrated manner or in a hierarchical manner in which elements are preferentially bound to the event's agent. In an integrated binding structure event elements are bound into a unitary representation or engram and can thus be accessed in an all-or-none manner (Damasio, 1989; Marr, 1971; Moll & Miikkulainen, 1997; Tulving, 1983; see also Shohamy & Wagner, 2008). A hierarchical binding structure consists of a system of pairwise bindings (see Cai et al., 2016; Cohen & Eichenbaum, 1993; Eichenbaum, 1999; Healy & Caudell, 2019; Hommel et al., 2001; Kumaran & McClelland, 2012; McClelland et al., 1995; Moeller et al., 2019), in which event elements are preferentially bound to particular types of elements, such that some bindings are systematically prioritized over others (see Schreiner et al., 2022). While initial findings (Horner et al., 2015; Horner & Burgess, 2014; see also Joensen et al., 2020) seem to favor an integrated binding structure, we previously found mixed evidence for the different structures, with some evidence favoring a hierarchical binding structure in which event elements are preferentially bound to the animate element and other evidence favoring an integrated binding structure (Schreiner et al., 2022).

Integrated and hierarchical binding structures make different predictions regarding the pattern of dependency across open-loop structures (see Horner et al., 2015; Horner & Burgess, 2014) in which different associations are excluded (see Schreiner et al., 2022). An integrated binding structure suggests that dependency does not vary as a function of the excluded association, because event elements are stored in a unitary representation that is accessible in an all-or-none manner. Thus, the association that is not presented should either be readily retrieved with the other associations (and thus dependency should be the same no matter which association is excluded) or the non-coherence of the encoding episodes may prevent integration and there should be no dependency in all open-loop conditions (and thus also no variation in dependency across open-loop conditions). A hierarchical binding structure suggests a system of pairwise bindings with asymmetrical binding strengths, with event elements being preferentially bound to particular types of elements. Thus, excluding a more critical association should more strongly diminish dependency than excluding a less critical association, and consequently dependency should vary as a function of the excluded association. We expected that event elements are preferentially bound to the agent of the event, and thus expected to find a stronger stochastic dependency of the retrieval of event elements when excluding associations not involving an agentic element than when excluding associations involving an agentic element in non-coherent encoding episodes (Hypothesis 2). For events without an agentic element we expected no differences in the stochastic dependency of the retrieval of event elements in non-coherent encoding episodes (Hypothesis 3), because in this case all associations should be equally important.

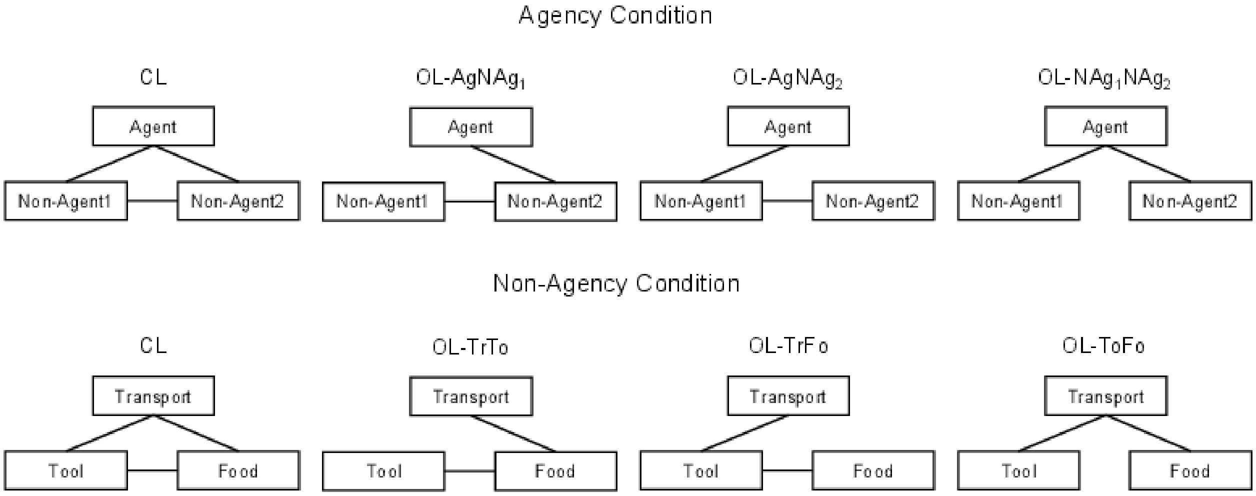
Procedure

In each of the three open-loop conditions we consistently excluded one pairwise association from presentation (see also the paired-associates learning paradigm, e.g., Preston et al., 2004). Thus, in the agency condition there was one condition in which the association agent – non-agent$_1$ was excluded (OL-AgNAg$_1$), one condition in which the association agent – non-agent$_2$ was excluded (OL-AgNAg$_2$), and one condition in which the association non-agent$_1$ – non-agent$_2$ was excluded (OL-NAg$_1$NAg$_2$). In the non-agency condition there was one condition in which the association means of transportation – tool was excluded (OL-TrTo), one condition in which the association means of transportation – food was excluded (OL-TrFo), and one condition in which the association tool – food was excluded (OL-ToFo, see Figure A.1). The open-loop conditions were equated to the closed-loop conditions regarding the number of event elements rather than the number of presented associations. Previous studies yielded similar results irrespective of whether conditions were equated regarding the number of event elements or the number of presented associations (Horner & Burgess, 2014; Joensen et al., 2020).

While event presentation in the closed-loop conditions consisted of three learning trials, event presentation in the open-loop conditions consisted of two learning trials. For the open-loop conditions, test trials included one trial per event with a cue-target pair that was not presented jointly in the learning phase (inference trials). However, the respective cue and target overlapped with a common event element and could thus be flexibly related to enable reconstruction of the association that was excluded from presentation in the learning phase. Loop conditions OL-AgNAg$_1$ and OL-TrTo, OL-NAg$_2$ and OL-TrFo, and OL-NAg$_1$NAg$_2$ and OL-ToFo were coerced into a common factor level, respectively, to jointly include them in the models for the analysis of memory performance.

Figure A.1

Associative Structure of the Experimental Conditions in Experiment 2



Note. CL = closed-loop, OL-AgNAg$_1$ = open-loop with association agent – non-agent$_1$ excluded, OL-AgNAg$_2$ = open-loop with association agent – non-agent$_2$ excluded, OL-NAg$_1$NAg$_2$ = open-loop with association non-agent$_1$ – non-agent$_2$ excluded, OL-TrTo = open-loop with association means of transportation – tool excluded, OL-TrFo = open-loop with association means of transportation – food excluded, OL-ToFo = open-loop with association tool – food excluded, Transport = means of transportation.

Dependency Results for the Open-Loop Conditions

There was no significant dependency in any of the open-loop conditions. In the agency condition, dependencies were $D = 0.01$ ($SE = 0.02$, $p = .45$) in loop condition OL-AgNAg$_1$, $D =$ -0.02 ($SE = 0.02$, $p = .13$) in loop condition OL-AgNAg$_2$, and $D =$ -0.01 ($SE = 0.02$, $p = .49$) in loop condition OL-NAg$_1$NAg$_2$. In the non-agency condition, dependencies were $D = 0.02$ ($SE = 0.02$, $p = .29$) in loop condition OL-TrTo, $D = 0.02$ ($SE = 0.02$, $p = .20$) in loop condition OL-TrFo, and $D = 0.00$ ($SE = 0.02$, $p = .84$) in loop condition OL-ToFo. The only significant dependency differences were a smaller dependency in loop condition OL-AgNAg$_2$ in the agency condition than in loop conditions CL ($D_{\text{diff}} =$ -0.05, $SE = 0.02$, $p = .02$), OL-TrTo ($D_{\text{diff}} =$ -0.04, $SE = 0.02$, $p = .04$), and OL-TrFo ($D_{\text{diff}} =$ -0.05, $SE = 0.02$, $p = .03$) in the non-agency condition. All other dependency differences were non-significant ($p \geq .06$).

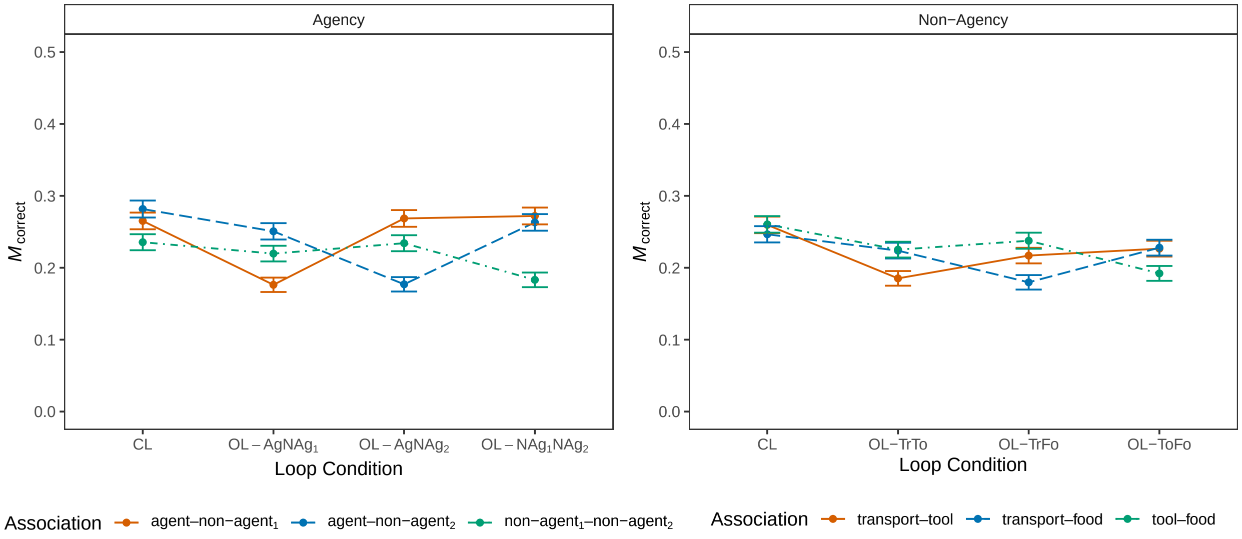
Given that there was no significant dependency in any of the experimental conditions, not even in the closed-loop conditions, as has been previously found (Bisby et al., 2018; Horner et al., 2015; Horner & Burgess, 2014; Joensen et al., 2020; Schreiner et al., 2022), the results cannot properly distinguish between an integrated and a hierarchical binding structure. Participants seem to have been unable to form coherent memory representations in all conditions of Experiment 2.

Appendix B

Figures and Tables for the Analysis of Memory Performance in Experiment 2

Figure B.1

Proportion of Correct Responses by Agency Condition, Loop Condition, and Association in Experiment 2



Notes. CL = closed-loop, OL-AgNAg$_1$ = open-loop with association agent – non-agent$_1$ excluded, OL-AgNAg$_2$ = open-loop with association agent – non-agent$_2$ excluded, OL-NAg$_1$NAg$_2$ = open-loop with association non-agent$_1$ – non-agent$_2$ excluded, OL-TrTo = open-loop with association means of transportation – tool excluded, OL-TrFo = open-loop with association means of transportation – food excluded, OL-ToFo = open-loop with association tool – food excluded, transport = means of transportation. Error bars represent ± *SEM*.

Table B.1

Results of the Post-Hoc Pairwise Comparisons for the Interaction of Agency Condition and Association Regarding Memory Performance in Experiment 2

Agency Condition	Contrast	*log OR*	95% CI
agency	agent–non-agent$_1$/transport–tool - agent–non-agent$_2$/transport–food	0.01	[-0.07, 0.10]
agency	agent–non-agent$_1$/transport–tool - non-agent$_1$–non-agent$_2$/tool–food	0.15	[0.06, 0.24]
agency	agent–non-agent$_2$/transport–food - non-agent$_1$–non-agent$_2$/tool–food	0.14	[0.05, 0.23]
non-agency	agent–non-agent$_1$/transport–tool - agent–non-agent$_2$/transport–food	0.02	[-0.07, 0.11]
non-agency	agent–non-agent$_1$/transport–tool - non-agent$_1$–non-agent$_2$/tool–food	-0.04	[-0.13, 0.05]
non-agency	agent–non-agent$_2$/transport–food - non-agent$_1$–non-agent$_2$/tool–food	-0.06	[-0.15, 0.03]

Note. log OR = log-odds ratio (median of the posterior distribution), CI = credible interval (highest posterior density interval), transport = means of transportation. Associations separated by a slash (/) were coerced into one factor level. Estimates are based on the full model.

Table B.2

Results of the Post-Hoc Pairwise Comparisons for the Interaction of Loop Condition and Association Regarding Memory Performance in Experiment 2

Loop Condition	Contrast	*log OR*	95% CI
CL	agent–non-agent$_1$/transport–tool - agent–non-agent$_2$/transport–food	-0.01	[-0.13, 0.11]
CL	agent–non-agent$_1$/transport–tool - non-agent$_1$–non-agent$_2$/tool–food	0.08	[-0.04, 0.20]
CL	agent–non-agent$_2$/transport–food - non-agent$_1$–non-agent$_2$/tool–food	0.09	[-0.03, 0.21]
OL-AgNAg$_1$/OL-TrTo	agent–non-agent$_1$/transport–tool - agent–non-agent$_2$/transport–food	-0.35	[-0.48, -0.22]
OL-AgNAg$_1$/OL-TrTo	agent–non-agent$_1$/transport–tool - non-agent$_1$–non-agent$_2$/tool–food	-0.27	[-0.40, -0.13]
OL-AgNAg$_1$/OL-TrTo	agent–non-agent$_2$/transport–food - non-agent$_1$–non-agent$_2$/tool–food	0.09	[-0.04, 0.21]
OL-AgNAg$_2$/OL-TrFo	agent–non-agent$_1$/transport–tool - agent–non-agent$_2$/transport–food	0.40	[0.27, 0.53]
OL-AgNAg$_2$/OL-TrFo	agent–non-agent$_1$/transport–tool - non-agent$_1$–non-agent$_2$/tool–food	0.03	[-0.09, 0.16]
OL-AgNAg$_2$/OL-TrFo	agent–non-agent$_2$/transport–food - non-agent$_1$–non-agent$_2$/tool–food	-0.36	[-0.49, -0.23]
OL-NAg$_1$NAg$_2$/OL-ToFo	agent–non-agent$_1$/transport–tool - agent–non-agent$_2$/transport–food	0.02	[-0.10, 0.14]
OL-NAg$_1$NAg$_2$/OL-ToFo	agent–non-agent$_1$/transport–tool - non-agent$_1$–non-agent$_2$/tool–food	0.37	[0.24, 0.50]
OL-NAg$_1$NAg$_2$/OL-ToFo	agent–non-agent$_2$/transport–food - non-agent$_1$–non-agent$_2$/tool–food	0.35	[0.23, 0.48]

Table B.2 continued

Loop Condition	Contrast	*log OR*	95% CI

Note. log OR = log-odds ratio (median of the posterior distribution), CI = credible interval (highest posterior density interval), CL = closed-loop, OL-AgNAg$_1$ = open-loop with association agent – non-agent$_1$ excluded, OL-AgNAg$_2$ = open-loop with association agent – non-agent$_2$ excluded, OL-NAg$_1$NAg$_2$ = open-loop with association non-agent$_1$ – non-agent$_2$ excluded, OL-TrTo = open-loop with association means of transportation – tool excluded, OL-TrFo = open-loop with association means of transportation – food excluded, OL-ToFo = open-loop with association tool – food excluded, transport = means of transportation. Loop conditions and associations separated by a slash (/) were coerced into one factor level. Estimates are based on the full model.