Beneficial and detrimental effects of schema incongruence on memory for contextual events

Darya Frank[1], Daniela Montaldi[1], Bianca Wittmann[2], Deborah Talmi[1]

[1] Division of Neuroscience and Experimental Psychology, University of Manchester, UK

[2] Department of Psychology, Justus Liebig University, Giessen, Germany

Word count: 5,560 words

Corresponding Author:

Darya Frank

Division of Neuroscience and Experimental Psychology

Zochonis Building

University of Manchester

Manchester, M13 9PL

Tel: +44 (0)161 2752690

Darya.frank@manchester.ac.uk

Abstract

Mental schemas provide a framework into which new information can easily be integrated. In a series of experiments, we examined how incongruence that stems from a prediction error (van Kesteren, Ruiter, Fernández, & Henson, 2012) modulates memory for multi-component events that instantiated pre-existing schemas. Each event consisted of four stimulus pairs with overlapping components, presented in four blocks (A-B, B-C, C-D, D-A). A-B pairs elicited contextual expectations (A: Farm, B: Tractor) that were either met by a congruent C component (C: Farmer) or violated by an incongruent one (C: Lawyer). The baseline condition included unrelated pairs, where the C component was neither congruent nor incongruent. In experiment 2, events were presented in successive trials instead of blocks, and eye movements were recorded to analyse allocation of attention. Memory was tested through old-new item recognition followed by cued recall. Across experiments, recognition and recall performance for incongruent components was reduced compared to congruent components. Incongruent items were in some cases more accurately retrieved compared to unrelated ones, depending on task demands. Additionally, better recall was observed in the incongruent D-A pairs, compared to congruent and unrelated ones, because of reduced interference from C components. Eye-tracking revealed an increased number of fixations on C components in the incongruent and unrelated conditions. These results suggest that the integration of incongruent items into an episode is impaired, compared to congruent items, despite the contextual surprise and increased attention they elicited at encoding. However, there was a beneficial effect of prediction error on memory performance, compared to a baseline, depending on the task employed.

Keywords: congruence effects, schema, prediction error, event integration

Abstract word count: 259 words

Introduction

Schemas and their role in supporting memory have been a topic of intensive research over several decades (Bartlett, 1932). Schema effects can benefit memory in two ways; firstly, the mere existence of a schema could enhance learning of new information. Previous research suggests memory performance is improved when the items can be assimilated into an existing schema (McClelland, 2013; Tse et al., 2007, 2011). Secondly, performance can be assessed across congruent and incongruent information. A common observation, referred to as the congruency effect, is that schema-congruent information is better recognised and recalled than incongruent information (Atienza, Crespo-Garcia, & Cantero, 2010; Craik & Tulving, 1975; Staresina, Gray, & Davachi, 2009; van Kesteren, Rijpkema, Ruiter, & Fernandez, 2010). However, van Kesteren and colleagues (2012) proposed a model (SLIMM) showing schemas can enhance memory for both congruent and incongruent information, via different mechanisms. The key modulator in this model is the prediction error elicited by the incongruent item (Greve, Cooper, Kaula, Anderson, & Henson, 2017). A prominent aspect that remains elusive is to what extent memory is enhanced by the prediction error (e.g. whether memory performance is equivalently good for congruent and incongruent items). Here, we test behavioural predictions of this model by combining contextual surprise with schema-incongruent items, embedded in multi-component events (Horner, Bisby, Bush, Lin, & Burgess, 2015; Horner & Burgess, 2013).

SLIMM posits that incongruence could lead to superior memory when the contextual schema provides a strong constraint (van Kesteren et al., 2012). In such cases, the incongruent item elicits a prediction error, leading to better memory through the creation of new representations. Importantly, the degree to which incongruence benefits memory, compared to a schema-less, baseline level, remains unclear. The model also accounts for congruency effects, as encountering congruent information results in medial prefrontal cortex (mPFC) activation of the schema, which in turn facilitates encoding in the medial temporal lobe (MTL; for similar ideas see Brod, Werkle-Bergner, & Shing, 2013; Preston & Eichenbaum, 2013). A key aspect of the SLIMM model is that the fate of incongruent information in memory is determined by the context it is embedded in. Most of the studies examining schema effects build upon pre-experimental knowledge (Bayen & Kuhlmann, 2011; Bein, Reggev, & Maril, 2014; van Kesteren et al., 2010) and the relationship between a pair of items, or their level of semantic relatedness (Bein et al., 2014; Staresina et al., 2009; van Kesteren et al., 2013). For example, purple-banana would be an incongruent pair, whereas yellow-banana constitutes a congruent one. However, this design does not necessarily allow for predictions to develop at encoding. Subsequently, during retrieval, there is no episodic contextual setting that would reinstate the schema (van Kesteren et al., 2012).

Incongruence that stems from a prediction error should be accompanied by better retention, supported by MTL engagement (Greve et al., 2017; van Kesteren et al., 2012). Importantly, for a prediction to be wrong, it must first be elicited (Kumaran & Maguire, 2007). We therefore utilised interleaved learning of events comprising pairs of components, to allow for predictions to be developed and violated. Previous research suggests such multi-component events are well suited for this purpose, as they promote binding of components into contextual events (Horner et al., 2015; Horner & Burgess, 2013). Therefore, the first pair of components presented can be used to implicitly set the schema-related predictions (Schlichting & Preston, 2015). Additionally, this paradigm makes it possible to test how incongruent components affect the rest of the contextual event. Specifically, whether incongruent components can be integrated into an otherwise congruent event (Bein et al., 2015), relying on interactions between mPFC and MTL (Schlichting & Preston, 2015, 2017). Thus, we can measure not only the independent recollection of incongruent components, but also their indirect effect on adjacent congruent components. According to SLIMM (van Kesteren et al., 2012), if incongruent representations are reactivated at retrieval (similarly to congruent ones), we would expect equivalent levels of interference from incongruent and congruent C components in adjacent A-B and D-A pairs.

In the studies reported here, we used events consisting of four pairs. Each pair shared a common component (A-B, B-C, C-D, D-A), to promote retrieval of the previous pair during study (Caplan, Rehani, & Andrews, 2014), as well as to allow for components to be integrated into an event (Burton, Lek, & Caplan, 2017; Schlichting & Preston, 2017). We extend previous literature on schema effects by actively eliciting contextual predictions that stem from existing schemas, as opposed to relying solely on the level of relatedness of two items. On this basis, we hypothesised that incongruence will modulate memory both of the incongruent components and the event they are incorporated in. Whilst SLIMM predicts memory enhancement of incongruent components (van Kesteren et al., 2012), it remains unclear to what extent. By including an unrelated 'baseline' condition, where there is no schema, we will test not only whether the presence of a prediction error supports schema-related memory, but also to what extent. For example, congruent and incongruent items could be equally better than unrelated ones, show a graded response (congruent > incongruent > unrelated), or be equivalent to unrelated events.

To assess differences in memory performance between congruent and incongruent items, compared to an unrelated baseline, we used contextual events (see Experiment 1a in the Supplementary Materials for comparison between congruent-incongruent alone). The first pair in the event is location-object, as it easily instantiates the schema (or lack thereof) for the following items (farm-tractor immediately brings to mind other farm-related items, whereas golf course-torch does not intrinsically belong to a specific context or schema, see

Bar & Aminoff, 2003). Both congruent and incongruent events had three components that were schema-congruent (A,B,D; see Figure 1 for examples). In the schema-based conditions, A-B pairs elicited contextual expectations that were either met by a congruent C component or violated by an incongruent one. In unrelated events, components did not share any common contextual information. Memory was tested in two steps, first a yes/no item recognition task for each component, followed by a cued recall for the adjacent component (e.g. B-? or ?-D, see Figure 1). Cued recall was tested only for initially recognised components. Multiple retrieval trials were employed to test effects of task demand (recognition vs. recall, forward vs. backward cued recall).

Figure 1 about here

Experiment 1

Results and Discussion

Item recognition. A three (congruence: congruent, incongruent, unrelated) by two (components: A, B, C, D) repeated measures ANOVA (Figure 2c) was conducted for corrected recognition responses (hits – false alarms). Greenhouse-Geisser correction for the sphericity assumption of ANOVAs are reported where appropriate, and all post-hoc analyses reported are Bonferroni corrected. When multiple t-tests were computed, a threshold p value is reported for all of them (e.g. all p's $< .05$). Despite near-ceiling recognition performance (overall average of 93% accuracy), a main effect of congruence was found $F(2,60) = 3.87$ $p = .026$, $\eta_p^2 = .113$, with follow-up paired t-tests indicating congruent components were more easily recognised than incongruent ones $t(30) = 2.95$, $p = .006$, Cohen's $d = .53$, and unrelated cues $t(30) = .258$, $p = .015$, Cohen's $d = .464$. No differences were observed between incongruent and unrelated components $t(30) = .851$, $p = .4$.

Cued recall. We first tested whether there were any effects of order on trials (AB - forward, BA - backward) from the same pair (A-B). The three (congruence) by four (pair) by two (order) repeated measures ANOVA yielded both a main effect of order $F(1,30) = 11.2$, $p = .002$, $\eta_p^2 = .272$ and a significant three-way interaction $F(6,180) = 4.2$, $p = .008$, $\eta_p^2 = .123$. This suggests there were different patterns of performance in each condition between the forward and backwards trials (see Figures 2a and 2b). Therefore, we carried out two separate three (congruence) by four (pair) ANOVAs, one for each order (forward X – ?, and backward ? – X). The forward ANOVA (Figure 2a) revealed a significant interaction between congruence and pairs $F(6,180) = 5.37$, $p < .001$, $\eta_p^2 = .152$. Similarly, the ANOVA for backward pairs (Figure 4b) revealed an interaction effect $F(6,180) = 4.01$, $p < .001$, $\eta_p^2 = .118$.

Planned comparisons between the congruence conditions for each trial were carried out to examine differences between conditions. Congruent BC and CD trials (forward order) were better recalled than incongruent and unrelated ones (all p's ≤ .011). No significant differences were observed between incongruent and unrelated components in these two trials (all p's ≥ .32). In the backward order trials, a benefit of incongruent over unrelated trials was found for DC $t(30) = 2.774$, $p = .009$, Cohen's $d = .5$, and a trend towards it was observed in CB trials $t(30) = 1.75$, $p = .091$, Cohen's $d = .31$. These results show the order effect was due to a graded pattern (congruent > incongruent > unrelated) in the backward, but not forward order. Finally, in DA and AD trials, incongruent components were better recalled than unrelated and congruent ones and incongruent AB were more accurately retrieved compared to congruent ones (all p's ≤ .01).

Figure 2 about here

Interference analysis. To further elucidate the benefit of incongruence on D-A and A-B pairs, we examined the erroneous answers for cued recall trials (Figure 2d). We inspected both trials comprising the D-A pairing (AD and DA) together, to have a sufficient number of trials and participants included (due to order effects in previous analysis we also carried out this analysis separated by order, which showed similar results despite a lower number of trials and participants included, see Supplementary Materials). We examined how many of the errors were due to interference from C items, compared to a baseline (erroneous recall of B/D). One out of the 31 participants had missing values for one pair, therefore data from 30 participants were analysed. A three (congruence: congruent, incongruent and unrelated) by two (item: C and D/B) by two (pair: A-B and D-A) repeated measures ANOVA revealed interaction effects of congruence by item $F(2,58) = 7.37$, $p = .001$, $\eta_p^2 = .203$, and item by pair $F(1,29) = 10.36$, $p = .003$, $\eta_p^2 = .263$. The three-way interaction effect was not significant $F(2,58) = 1.1$, $p = .34$.

Post-hoc tests revealed more C errors in congruent D-A and A-B pairs compared to incongruent and unrelated ones (all p's < .004). There were also more C errors in incongruent D-A and A-B pairs compared to unrelated ones (all p's < .038). When comparing interference between C items and baseline B/D items, in D-A pairs we observed less interference from incongruent C compared to incongruent B components ($t(29) = 2.14$, $p = .041$, Cohen's d = .391). In A-B trials, on the other hand, interference to incongruent pairs is equivalent between C and D items ($t(29) = .084$, $p = .934$), but there is more interference from congruent C items compared to congruent D ones ($t(29) = 2.43$, $p = .021$, Cohen's d = .444).

The results described above show that memory performance for congruent components is superior to incongruent and unrelated ones, in all testing formats. Incongruent items are, in backward cued recall, more accurately retrieved than unrelated ones. The interference analysis showed better recall performance for congruence-matched pairs (A-B, D-A) of incongruent events is due to reduced interference from C components, suggesting they are less integrated into the event. Additionally, incongruent C items caused more interference than unrelated ones, showing a similar pattern of responses as in backward cued recall. Overall, these findings suggest that congruence benefits memory performance across the board, but memory for incongruent components was modulated by task demands. Differences in performance between incongruent and unrelated pairs, stemming from the prediction error associated with incongruence, were observed in backward cued recall and the interference analysis. In both cases, a graded response was observed (congruent > incongruent > unrelated). In the recognition and forward recall, on the other hand, memory for incongruent components was equivalent to unrelated ones.

Experiment 2

To address the integration account of reduced performance for incongruent components, we devised Experiment 2 to allow for easier integration of the pairs into a cohesive event. To do so, events were presented as trial-by-trial pairs, rather than across blocks. Trial-by-trial presentation differs in the contextual setting of learning. In Experiment 1, each pair was compared to other pairs of the same kind (i.e. B-C pairs from different events were always studied together). In Experiment 2, on the other hand, the ‘reference point’ is the previous pair from the same event (B-C will follow A-B from the same event). This change would allow for a quicker build-up of predictions, as the event pairs would now be temporally closer than in the previous experiments. We added an associative inference task between A and C items, which were not shown together, to test differences in integration levels. To examine whether unexpected pairs are processed differently at encoding, we measured eye-movements during this stage. We reasoned that changes in fixation patterns observed at encoding would indicate a different allocation of resources to components that are more difficult to encode and integrate into the event.

Eye-tracking results

We first subjected the number of fixations to a three (congruence) by four (pair) by 2 (area of interest, AOI) ANOVA (Figure 3a), which showed a significant main effect of pair $F(2.1,56.73) = 73.6$, $p < .001$, $\eta_p^2 = .732$, a congruence by pair interaction $F(6,162) = 5.63$, $p <.001$, $\eta_p^2 = .173$ and a significant AOI by pair interaction $F(2.23,60.32) = 25.7$, $p < .001$, $\eta_p^2 = .49$. Post-hoc comparisons showed the least amount of fixations on the B item of unrelated A-B pairs, when compared to the congruent and incongruent conditions (all p's ≤ .008). Additionally, unrelated B-C pairs were associated with fewer fixations on B, when compared to congruent and incongruent pairs (all p's < .001). For the C components in the B-C pairs, congruent components were associated with the lowest number of fixations compared to incongruent and unrelated ones (all p's ≤ .008). A similar analysis for fixation durations (Figure 3b) revealed significant main effects of pair $F(1.53, 42.91) = 76.33$, $p < .001$, $\eta_p^2 = .732$ and AOI $F(1,28) = 13.21$, $p = .001$, $\eta_p^2 = .321$, as well as a pair by congruence interaction $F(6,168) = 5.9$, $p < .001$, $\eta_p^2 = .174$, and a pair by AOI interaction $F(3.84) = 19.98$, $p < .001$, $\eta_p^2 = .416$. Planned comparisons indicated that in unrelated B-C pairs, fixations on the B components were shortest and those on C components were longest, compared to congruent and incongruent pairs (all p's ≤ .003).

Figure 3 about here

Behavioural Results

Item recognition. A three (congruence: congruent, incongruent, unrelated) by four (component: A, B, C, D) repeated measures ANOVA was conducted for corrected recognition responses. Again, a main effect of congruence was found $F(2,54) = 14.83$ $p < .001$, $\eta_p^2 = .355$, with follow-up paired t-tests indicating congruent components were more easily recognised than incongruent and unrelated ones (all p's ≤ .002). No differences were observed between incongruent and unrelated components ($p > .2$).

Cued recall. We first tested whether there were any effects of order. Similar to Experiment 1, the three (congruence) by four (pair) by two (order) repeated measures ANOVA yielded a significant three-way interaction $F(6,162) = 4.39$, $p < .001$, $\eta_p^2 =.14$. This suggests there were different patterns of performance in each condition between the forward and backwards orders. Therefore, we performed two separate three (congruence) by four (pair) ANOVAs, one for each presentation order (forward X – ?, and backward ? – X). The forward ANOVA (Figure 4a) revealed a significant congruence by pair interaction $F(4.14, 112) = 5.67$, $p < .001$, $\eta_p^2 = .174$. The ANOVA for backward trials (Figure 4b) was akin to that for the forward trials, with an interaction between congruence and pair $F(6,162) = 9.37$, $p < .001$, $\eta_p^2= .258$.

Planned comparisons between the congruence conditions for each trial were then carried out. For AB and DA trials a similar pattern emerged, whereby there were no significant differences between congruent trials and any of the other conditions (all p's ≥ .02), but a significant benefit of incongruent trials over unrelated ones was observed (all p's ≤ .002). Unrelated BA trials were associated with reduced recall performance compared to congruent and incongruent ones (all p's ≤ .006). For BC, CB, CD and DC trials there were again similar findings, with congruent trials associated with better recall compared to incongruent and unrelated trials (all p's ≤ .013), but no difference was observed between incongruent and unrelated trials (all p's ≥ .668). These results suggest that performance for trials that are part of incongruent events tracks their level of relatedness.

Figure 4 about here

Associative inference retrieval task (A-C). A three (congruence) by two (component: A vs. C) ANOVA yielded a significant main effect of congruence $F(1.57, 40.97) = 105.46$, $p < .001$, $\eta_p^2 = .802$ (Figure 4c). The main effect of component was not significant $F(1,26) = .016$, $p = .901$, as was the interaction effect $F(1.43, 37.23) = .58$, $p = .508$. Post-hoc tests revealed that congruent components were associated with better performance compared to incongruent and unrelated ones, and a benefit of incongruent components over unrelated ones (all p's < .001).

Interference analysis. . A three (congruence: congruent, incongruent and unrelated) by two (component: C and D/B) by two (pair: A-B and D-A) repeated measures ANOVA (Figure 4d) revealed a significant interaction between congruence and item $F(2,54) = 34.8$, $p < .001$, $\eta_p^2 = .563$). The other two-way interactions, as well the three-way interaction were not significant (all p's > .248). Planned comparisons showed increased interference from congruent C components in A-B and D-A pairs, compared to incongruent and unrelated components (all p's < .001). There were no significant differences between incongruent and unrelated C components (all p's > .528). Additionally, there was more interference from incongruent B/D (baseline) components, compared to incongruent C components, in D-A pairs ($t(27) = 4.45$, $p < .001$, Cohen's $d = .841$) and A-B pairs, respectively ($t(27) = 3.44$, $p = .002$, Cohen's $d = .778$).

General Discussion

In a series of experiments, we found that unexpected incongruent components were less likely to be recognised and recalled, compared to congruent ones. This pattern was observed both when events were presented across blocks, to allow for predictions to develop gradually, and when events were constructed trial-by-trial, to promote easier integration. Examination of fixation patterns during encoding of such events revealed increased fixations on the first unexpected incongruent component, compared to their congruent counterparts, suggesting they were more difficult to encode. Our results suggest the presence of a prediction error in incongruent pairs did not enhance memory to the level observed in congruent events. However, depending on task demands, it did benefit memory compared to unrelated components (showing a graded pattern of responses). Finally, a beneficial effect of incongruence was observed in A-B and D-A pairs, which were only indirectly related to the incongruent C component.

The advantageous role of schema congruence on memory performance, compared to both incongruent and unrelated items, is in line with the congruency effect. Previous findings, and predictions from SLIMM, suggest congruent items benefit from the existing strong representation of the schema (Craik & Tulving, 1975; Staresina et al., 2009; van Kesteren et al., 2012). Previous experiences of similar associations (e.g. farm-tractor) are reactivated by the mPFC (Brod et al., 2013; Preston & Eichenbaum, 2013; Schlichting & Preston, 2015), and proposed to be more readily available during retrieval (Hemmer & Steyvers, 2009; Moscovitch & Craik, 1976; Steyvers & Hemmer, 2012). This notion is in line with our finding that congruent A-B and D-A pairs are more prone to interference from C components, showing reduced recall accuracy. Incongruent components, and unrelated ones even more so, are more difficult to integrate into the event (Bein et al., 2015; Craik & Tulving, 1975), resulting in less associative competition (Caplan et al., 2014), or a more constrained search space (Anderson, 1981) for retrieving A-B and D-A.

Further evidence for the reduced integration of incongruent C components can be found in forward-order trials in Experiment 1, as well as in all retrieval trials in Experiment 2. Incongruent components show comparable results, in memory performance and eye-fixation patterns, to those of their unrelated counterparts. The eye-tracking results indicate increased effort invested in encoding incongruent items, but this effort does not come into fruition later in retrieval performance. The encoding part of this effect can be attributed to MTL-driven function, showing the prediction error associated with an incongruent component promotes more elaborate encoding. During retrieval, congruent information seems to be dominating, especially in trials where a congruent item (B or D) cues retrieval of an incongruent item. In this case, mPFC involvement would potentially direct retrieval towards congruent representations (Preston & Eichenbaum, 2013; van Kesteren et al., 2012). Although previous studies have shown similar encoding and retrieval effects independently (van Kestern et al., 2010, 2013), future neuroimaging studies could examine the co-occurrence of such effects. Additionally, due to the use of bidirectional cued-recall tests, we could not directly correlate fixations at encoding with later retrieval performance, but this would be an interesting effect to examine.

Incongruent components also served as worse cues for their associates (Schulman, 1974), showing impaired recall performance (equivalent to unrelated pairs). This finding is more difficult to interpret in light of SLIMM's predictions (van Kesteren et al., 2012), as successful retrieval in this case requires reactivation of a congruent item. We argue this finding points to difficulty in binding the incongruent component into the event (Craik & Tulving, 1975; Bein et al., 2015; Packard et al., 2017), as also indicated by increased fixations. Given that schemas facilitate gist extraction and abstraction of commonalities (Gilboa & Marlatte, 2017), the presence of an incongruent component in the event could interrupt this process. This would therefore result in reduced schema instantiation to support binding of the incongruent C component to its congruent pairwise associates, explaining impaired performance on C-B and C-D pairs.

Interestingly, in the backward order retrieval trials and interference analysis (Experiment 1), as well as in the A-C inference task (Experiment 2), a graded pattern of responses was observed (congruent > incongruent > unrelated). To our knowledge, this is the first study to show such pattern. Although van Kesteren and colleagues (2012) postulated memory for both congruent and incongruent items would be enhanced, the extent to which this effect varies between conditions was unclear. Previous findings pertaining to congruency effects have not used an unrelated baseline condition, making it difficult to account for such differences. Here, we find that whilst components from congruent events were unequivocally better recalled, incongruent components showed better performance compared to unrelated ones. This pattern of results suggests a prediction error can enhance memory performance (compared to unrelated items), though not to the same extent as congruence does. Interestingly, this result was observed only under specific circumstances, suggesting this effect could be susceptible to task demands (Ghosh & Gilboa, 2014).

In Experiment 1, events were created across blocks, thus online comparisons were between B-C pairs from different events. Conversely, in Experiment 2 comparisons were made with A-B pairs from the same event. This difference in temporal context during encoding, could have biased processing of incongruent pairs in Experiment 1 compared to Experiment 2 (Howard & Kahana, 2002). Although overall memory performance in Experiment 1 was symmetric between forward and backward trials, graded responses were observed only in backward trials. Asymmetry in memory recall has been suggested to depend on the relationship between paired associates (Greene & Tussing, 2001; Li & Lewandowsky, 1995; Yang et al., 2013) and to engage the anterior hippocampus (Giovanello, Schnyer, &

Verfaellie, 2009). Forward recall is believed to be schema-driven (Geiselman & Callot, 1990) and more susceptible to disruptions during formation of associations at encoding (Li & Lewandowsky, 1995). In line with this, we find superior memory for congruent components, with no differences between incongruent and unrelated pairs in the forward order. Backward recall, on the other hand, is suggested to be more data-driven (Geiselman & Callot, 1990) and thus more susceptible to contextual details at encoding. We therefore suggest backward retrieval in our task promoted the beneficial effect of prediction error, mediated by hippocampal engagement (van Kesteren et al., 2012).

Graded responses were also obtained in the associative inference task in Experiment 2. Successful performance on such tasks is often used as a marker for schematic organisation of representations in memory, as it supports novel integration of indirectly related items (Kumaran, Summerfield, Hassabis, & Maguire, 2009; Preston & Eichenbaum, 2013; Tse et al., 2007; Zeithamova, Dominick, & Preston, 2012). In this task, incongruent and unrelated pairs were matched on relatedness and differed only on the build-up of expectations from the event's schema. Thus, the presence of a prediction error here could have mediated enhanced inference in this task. Alternatively, the mere existence of a schema in the incongruent condition, as opposed to the unrelated one, could have supported this inference (Kumaran, 2013; McClelland, 2013; Preston & Eichenbaum, 2013; Tse et al., 2011; Zeithamova et al., 2012). Future research on schema effects would benefit from further exploring these effects, specifically in relation to how task demands can modulate memory for incongruent information.

Our main aim was to test behavioural predictions outlined by SLIMM (van Kesteren et al., 2012). We thus utilised a paradigm that allows expectations to gradually develop by using interleaved learning of paired associates. The findings reported above provide some support to the notion prediction errors can enhance memory for incongruent items. An alternative interpretation is that the amount of prediction error associated with incongruent components in our studies was not large enough to result in conclusively improved performance. Critically, our task was designed to implicitly set participants' predictions. This was done in order to capture the inherent aspect of predictions as they arise in daily life, and to avoid any artificial allocation of attention towards this manipulation. As a result of this manipulation, we could not quantify the amount of prediction error elicited by incongruent components, but only indirectly assert contextual predictions were elicited by the stimuli used (Bar & Aminoff, 2003).

In conclusion, our results provide further evidence for the notion that schemas aid memory by providing a structured representation into which congruent information can easily fit. The findings reported here also shed light on the extent to which prediction errors in incongruent items support its presence in memory. Although it requires more effortful encoding, retrieval success of incongruent items is always reduced compared to congruent components. The extent to which incongruent items are better remembered compared to unrelated components, on the other hand, is modulated by task demands. Future research looking into schema-mediated memory may build on the approach and findings highlighted above to better understand factors contributing to these effects.

Methods

Experiment 1

Participants. 35 participants (12 males) gave informed consent to take part in the experiment. Four participants whose recognition performance was either above or below three times the IQR were excluded from any further analysis. Thus, data from 31 participants between the ages 18-27 (M = 19.8, SD = 2.91) were analysed.

Materials. The experiment was controlled using E-Prime 1 (Psychology Software Tools). Stimuli were 30 four-components events (10 congruent, 10 incongruent 10 unrelated events). Each event contained a location (component A, e.g. farm), two objects (components B and D, haystack and a tractor) and a person's profession (component C, farmer). Items were presented as images with labels above them (see Figure 1 for examples). Congruent and incongruent events were constructed to elicit strong contextual predictions, meaning that their components are most likely to appear in the given context, as established by previous work (Bar & Aminoff, 2003). Ten of the events were assigned to be in the incongruent condition, such that the person (C component) was unexpected in the context (a lawyer in the context of a farm with a haystack and a tractor). Another 10 were congruent, meaning the person was expected given the context (a farmer in a farm). The final 10 events included objects with low contextual value (Bar & Aminoff, 2003), meaning they can be found in in a variety of contexts. The allocation of events to conditions was counterbalanced across participants. Images were obtained from freely available online resources labelled with a Creative Commons License.

Procedure. The experiment consisted of three phases: encoding, distraction and retrieval (see Figure 1). The encoding phase was interleaved and took place over 4 blocks, one block for each pairwise association (A-B, B-C, C-D, D-A). Critically, A-B and D-A pairs are not directly associated with the C component (which defines whether an event is assigned to a congruent or incongruent condition). These pairs are congruent in both congruent and incongruent conditions, thus providing an opportunity for testing the effect of incongruence on the rest of the event. Each block consisted of 30 randomised-order trials, resulting in a total of 120 trials in the encoding phase. Each trial began with a one-second fixation cross, followed by presentation of one pair of labelled images for three seconds. Participants were instructed to imagine the components interacting together, as vividly as they could, while being aware of their respective location on the screen (right and left).

To prevent participants from actively rehearsing the encoded information, a distractor task involving solving arithmetic problems was used for five minutes. Participants were instructed to be as accurate as possible and were informed that if they failed to reach a certain performance threshold their data would be excluded from further analysis. Following this task, the retrieval phase began, where items were presented in a pseudo-randomised order, based on 10 pre-made lists. Two retrieval tasks were employed, a recognition task for each component, followed by a cued-recall task only for the recognised components.

Participants were first presented with the yes/no recognition task. They had a maximum of 10 seconds to complete this task. If they responded 'yes', a second cued-recall task took place immediately. For this task, the recognised component was coded as 'cue' and the recalled component as 'target'. Participants were asked to recall which item appeared with the previously recognised item (a source recall task, retrieve the item in the location indicated on the screen). Each pairwise association from each event was tested in both directions in a randomised order (for example, forward A-? and backward ?-B). A cued-recall answer was scored as correct if it was identical to the item presented at encoding, if it was a specific case of the same category (for example, 'car mechanic' instead of 'mechanic') or semantically similar ('gymnast' instead of 'acrobat'). If the participant failed to give an answer within 20 seconds of the cue being presented, the trial was scored as incorrect. In addition to the 120 items that were encoded, 30 additional items, making up six events, were used as foils.

Experiment 2

Experiment 2 was identical to Experiment 1, with the following modifications:

Participants. 35 participants (9 males) gave informed consent to take part in the experiment. Data from seven participants were removed from any further analysis due to technical failure of the eye-tracker (one participant), poor eye-tracking data (three participants; inclusion of these participants in the behavioural analysis did not change the results) and poor memory performance, above or below three times the IQR (three participants). Thus, data from 28 participants between the ages 18-28 (M = 20.8, SD = 2.78) were analysed.

Materials and apparatus. The materials were identical to those used in Experiment 1. To examine fixation patterns during encoding, eye movements were recorded during encoding, using an ASL infrared eye tracker (Eye-Trac 6000, Applied Science Laboratories) at a sampling rate of 60 Hz. The desktop-mounted camera was placed under the presentation screen, 70cm away from the participant. A chin-rest was used to minimise participants' movement.

Procedure. Before the experiment started, eye calibration was performed using a 9-point matrix. During the encoding phase in Experiment 2, events were presented sequentially, trial–by–trial, rather than across blocks as was done in Experiment 1. Each of the four pairs comprising an event was presented for three seconds, with a 1s fixation cross between them. Following the last pair of the event, a fixation cross was displayed for 2s, until a new event had started. To examine whether reduced performance for the incongruent C components was due to a difficulty in integrating them as part of the events, an additional retrieval inference task was employed. Following the cued-recall task, participants were presented with a location or a person (A or C components) for a maximum of 10 seconds and were asked to recall its counterpart person or location from the same event, respectively.

Eye tracking analysis. Eyeneal software (Applied Science Laboratories) was used to convert the raw gaze coordinates to fixation points. The start of a fixation point was defined as six sequential gaze points with a standard deviation smaller than 0.5 visual degrees. The end of a fixation was marked when three consecutive gaze points were at least one visual degree away from the initial fixation location. The fixation points reported below are the average point of the start and end fixation locations. Two areas of interest (AOI) were defined in the Fixplot software (Applied Science Laboratory), one for each of the images displayed on the screen. Statistical analyses were conducted on the number of fixations and fixation duration for each AOI.

Acknowledgments. The authors thank L. Bailey for help with data collection for Experiment 2. D.F. is supported by a PDS award from the University of Manchester.

Author Contributions Statement:
D.F., D.M., B.W., and D.T. designed the experiment, D.F collected and analysed data, D.F., D.M., B.W. and D.T. wrote the manuscript.

Competing financial interest statement:
The author(s) declare no competing financial interests.

References

Anderson, J. R. (1981). Effects of prior knowledge on memory for new information. *Memory & Cognition*, *9*(3), 237–246. Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-0002794789&partnerID=40&md5=7587d2d916c8d019e0564dfc47322d62%5Cn/Users/Fernando/Documents/Project Based learning/PBL Articles/Anderson1981.pdf

Atienza, M., Crespo-Garcia, M., & Cantero, J. L. (2010). Semantic congruence enhances memory of episodic associations: role of theta oscillations. *Journal of Cognitive Neuroscience*, *23*(1), 75–90. http://doi.org/10.1162/jocn.2009.21358

Bar, M., & Aminoff, E. (2003). Cortical analysis of visual context. *Neuron*, *38*(2), 347–358. http://doi.org/10.1016/S0896-6273(03)00167-3

Bartlett, F. C. (1932). *Remembering: an experimental and social study*. Cambridge: Cambridge University.

Bayen, U. J., & Kuhlmann, B. G. (2011). Influences of source-item contingency and schematic knowledge on source monitoring: Tests of the probability-matching account. *Journal of Memory and Language*, *64*(1), 1–17. http://doi.org/10.1016/j.jml.2010.09.001

Bein, O., Livneh, N., Reggev, N., Gilead, M., Goshen-Gottstein, Y., & Maril, A. (2015). Delineating the Effect of Semantic Congruency on Episodic Memory: The Role of Integration and Relatedness. *Plos One*, *10*(2). http://doi.org/10.1371/journal.pone.0115624

Bein, O., Reggev, N., & Maril, A. (2014). Prior knowledge influences on hippocampus and medial prefrontal cortex interactions in subsequent memory. *Neuropsychologia*, *64*(October 2015), 320–330. http://doi.org/10.1016/j.neuropsychologia.2014.09.046

Brod, G., Werkle-Bergner, M., & Shing, Y. L. (2013). The Influence of Prior Knowledge on Memory: A Developmental Cognitive Neuroscience Perspective. *Frontiers in Behavioral Neuroscience*, *7*(October), 1–13. http://doi.org/10.3389/fnbeh.2013.00139

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. *The Quarterly Journal of Experimental Psychology*, *70*(4), 832–857. http://doi.org/10.1080/17470218.2016.1171886

Caplan, J. B., Rehani, M., & Andrews, J. C. (2014). Associations compete directly in memory. *Quarterly Journal of Experimental Psychology (2006)*, *67*(5), 955–78. http://doi.org/10.1080/17470218.2013.838591

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, *104*(3), 268–294. http://doi.org/10.1037/0096-3445.104.3.268

Geiselman, R. E., & Callot, R. (1990). Reverse versus forward recall of script-based texts. *Applied Cognitive Psychology*, *4*(2), 141–144. http://doi.org/10.1002/acp.2350040206

Ghosh, V. E., & Gilboa, A. (2014). What is a memory schema? A historical perspective on current neuroscience literature. *Neuropsychologia*, *53*, 104–14. http://doi.org/10.1016/j.neuropsychologia.2013.11.010

Gilboa, A., & Marlatte, H. (2017). Neurobiology of Schemas and Schema-Mediated Memory. *Trends in Cognitive Sciences*, 1–14. http://doi.org/10.1016/j.tics.2017.04.013

Giovanello, K. S., Schnyer, D., & Verfaellie, M. (2009). Distinct hippocampal regions make unique contributions to relational memory. *Hippocampus*, *19*(2), 111–117. http://doi.org/10.1002/hipo.20491

Greene, R. L., & Tussing, A. A. (2001). Similarity and associative recognition. *Journal of Memory and Language*, *45*(4), 573–584. http://doi.org/10.1006/jmla.2001.2795

Greve, A., Cooper, E., Kaula, A., Anderson, M. C., & Henson, R. N. (2017). Does prediction error drive one-shot declarative learning ? *Journal of Memory and Language*, *94*, 149–165. http://doi.org/10.1016/j.jml.2016.11.001

Hemmer, P., & Steyvers, M. (2009). Integrating Episodic and Semantic Information in Memory for Natural Scenes. *Proceedings of the 31st Annual Conference of the Cognitive Science Society*, 1557–1562. Retrieved from http://csjarchive.cogsci.rpi.edu/proceedings/2009/papers/337/paper337.pdf

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*(7462), 7462. http://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), 1370–83. http://doi.org/10.1037/a0033626

Howard, M. W., & Kahana, M. J. (2002). When does semantic similarity help episodic retrieval? *Journal of Memory and Language*, *46*(1), 85–98. http://doi.org/10.1006/jmla.2001.2798

Kahana, M. J., & Caplan, J. B. (2002). Associative asymmetry in probed recall of serial lists. *Memory and Cognition*, *30*(6), 841–849. http://doi.org/10.3758/BF03195770

Kumaran, D. (2013). Schema-driven facilitation of new hierarchy learning in the transitive inference paradigm. *Learning & Memory*, 388–394.

Kumaran, D., & Maguire, E. A. (2007). Which Computational Mechanisms Operate in the Hippocampus During Novelty Detection ?, *748*, 735–748. http://doi.org/10.1002/hipo

Kumaran, D., Summerfield, J. J., Hassabis, D., & Maguire, E. a. (2009). Tracking the Emergence of Conceptual Knowledge during Human Decision Making. *Neuron*, *63*(6), 889–901. http://doi.org/10.1016/j.neuron.2009.07.030

Li, S.-C., & Lewandowsky, S. (1995). Forward and Backward Recall: Different Retrieval Processes. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *21*(4), 837–847.

McClelland, J. L. (2013). Incorporating rapid neocortical learning of new schema-consistent information into complementary learning systems theory. *Journal of Experimental Psychology: General*, *142*(4), 1190–1210. http://doi.org/10.1037/a0033812

Moscovitch, M., & Craik, F. I. M. (1976). Depth of processing, retrieval cuesm and uniqueness of encoding as factors in recall. *Journal of Verbal Learning and Verbal Behavior*, *15*, 447–458.

Packard, P. A., Rodríguez-Fornells, A., Bunzeck, N., Nicolás, B., de Diego-Balaguer, R., & Fuentemilla, L. (2017). Semantic Congruence Accelerates the Onset of the Neural Signals of Successful Memory Encoding. *The Journal of Neuroscience*, *37*(2), 291–301. http://doi.org/10.1523/JNEUROSCI.1622-16.2017

Preston, A. R., & Eichenbaum, H. (2013). Interplay of Hippocampus and Prefrontal Cortex in Memory. *Current Biology*, *23*(17), R764–R773. http://doi.org/10.1016/j.cub.2013.05.041

Schlichting, M. L., & Preston, A. R. (2015). Memory integration: neural mechanisms and implications for behavior. *Current Opinion in Behavioral Sciences*, *1*, 1–8. http://doi.org/10.1016/j.cobeha.2014.07.005

Schlichting, M. L., & Preston, A. R. (2017). The Hippocampus and Memory Integration: Building Knowledge to Navigate Future Decisions. In D. E. Hannula & M. C. Duff (Eds.), *The Hippocampus from Cells to Systems: Structure, Connectivity, and Functional Contributions to Memory and Flexible Cognition* (pp. 405–437). Cham: Springer International Publishing. http://doi.org/10.1007/978-3-319-50406-3_13

Schulman, A. I. (1974). Memory for words recently classified. *Memory & Cognition*, *2*(1), 47–52. http://doi.org/10.3758/BF03197491

Staresina, B. P., Gray, J. C., & Davachi, L. (2009). Event Congruency Enhances Episodic Memory Encoding through Semantic Elaboration and Relational Binding. *Cerebral Cortex*, *19*(5), 1198–1207. http://doi.org/10.1093/cercor/bhn165

Steyvers, M., & Hemmer, P. (2012). Reconstruction from Memory in Naturalistic Environments. In B. H. Ross (Ed.), *The Psychology of Learning and Motivation* (Vol. 56, pp. 125–144). http://doi.org/10.1016/B978-0-12-394393-4.00004-2

Tse, D., Langston, R. F., Kakeyama, M., Bethus, I., Spooner, P. A., Wood, E. R., … Morris, R. G. M. (2007). Schemas and Memory Consolidation. *Science*, *316*(5821), 76–82. http://doi.org/10.1126/science.1135935

Tse, D., Takeuchi, T., Kakeyama, M., Kajii, Y., Okuno, H., Tohyama, C., … Morris, R. G. M. M. (2011). Schema-Dependent Gene Activation. *Science*, *333*(August), 891–895. http://doi.org/10.1126/science.1205274

van Kesteren, M. T. R., Beul, S. F., Takashima, A., Henson, R. N., Ruiter, D. J., & Fernández, G. (2013). Differential roles for medial prefrontal and medial temporal cortices in schema-dependent encoding: From congruent to incongruent. *Neuropsychologia*, *51*(12), 2352–2359. http://doi.org/10.1016/j.neuropsychologia.2013.05.027

van Kesteren, M. T. R., Rijpkema, M., Ruiter, D. J., & Fernandez, G. (2010). Retrieval of Associative Information Congruent with Prior Knowledge Is Related to Increased Medial Prefrontal Activity and Connectivity. *Journal of Neuroscience*, *30*(47), 15888–15894. http://doi.org/10.1523/JNEUROSCI.2674-10.2010

van Kesteren, M. T. R., Ruiter, D. J., Fernández, G., & Henson, R. N. (2012). How schema and novelty augment memory formation. *Trends in Neurosciences*, *35*(4), 211–219. http://doi.org/10.1016/j.tins.2012.02.001

Yang, J., Zhao, P., Zhu, Z., Mecklinger, A., Fang, Z., & Li, H. (2013). Memory asymmetry of forward and backward associations in recognition tasks. *Journal of Experimental Psychology: Learning Memory and Cognition*, *39*(1), 253–269. http://doi.org/10.1037/a0028875

Zeithamova, D., Dominick, A. L., & Preston, A. R. (2012). Hippocampal and Ventral Medial Prefrontal Activation during Retrieval-Mediated Learning Supports Novel Inference. *Neuron*, *75*(1), 168–179. http://doi.org/10.1016/j.neuron.2012.05.010

Figures

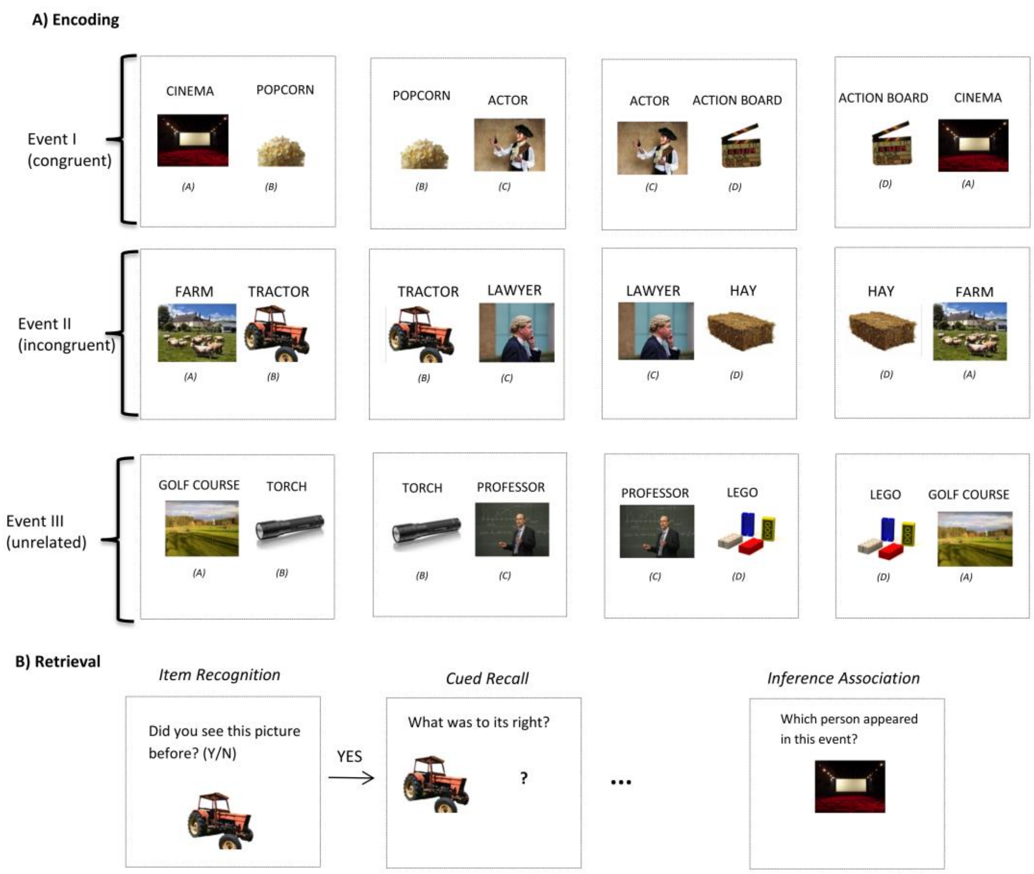

Figure 1 | Experimental Design. A) Encoding phase. Participants 120 paired associates, over four blocks, one for each pairwise association from every event. They imagined each pair interacting in a meaningful way for 3s. Each pair was preceded by a 1s fixation cross. B) Retrieval phase. Participants were presented with a cue and asked to indicate whether they remember seeing it at encoding. If they responded 'yes' they were asked to recall one of the other components from the same event, based on the spatial location of the cue. Inference association task was used in Experiment 2. Labels in parentheses are for illustration and were not presented during the experiment.

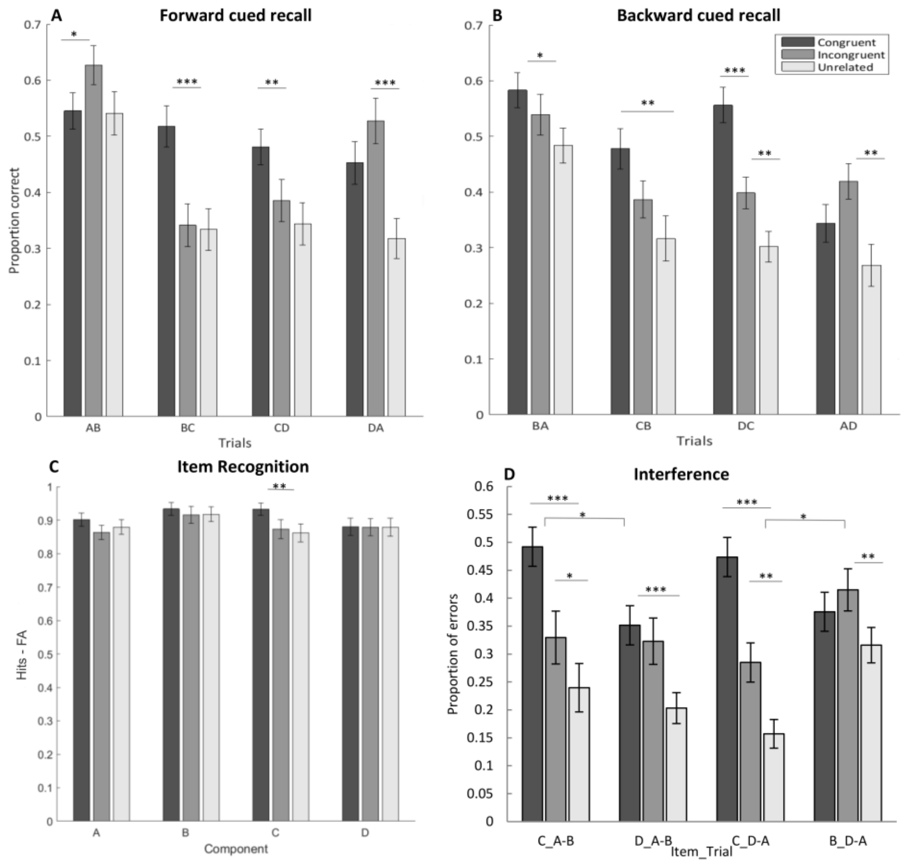

Figure 2 | Results Experiment 1. A) Cued recall performance forward pairs (e.g. A-?). B) Cued recall performance backward pairs (e.g. ?-B). In both orders performance tracks levels of relatedness of pairs, such that most incongruent and unrelated pairs are equivocal. C) Item recognition, Congruent C components show better accuracy compared to incongruent ones. D) Interference analysis, percentage of erroneously recalled C items in the cued recall task. Most interference from congruent items, followed by incongruent and then unrelated. Unless otherwise states, error bars represent standard error of mean. * p ≤ .05, ** p ≤ .01, *** p ≤ .001

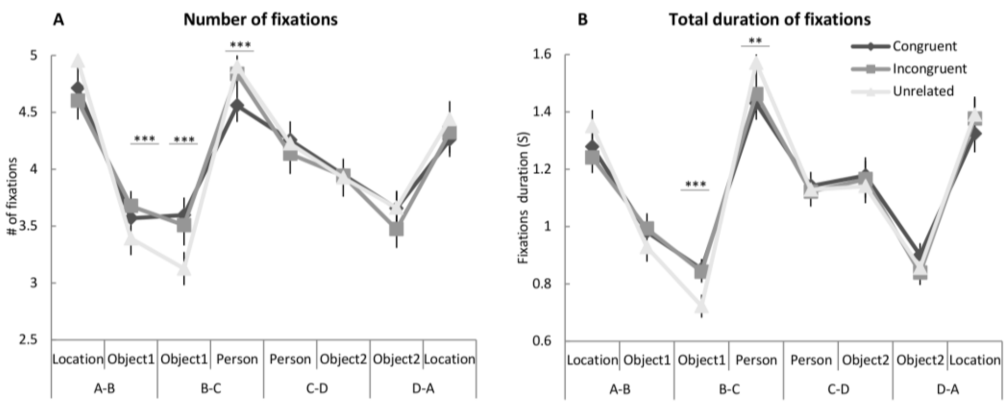

Figure 3 | Eye tracking results Experiment 2. A) Number of fixations on each item, per pair. Increased fixations on first presentation of incongruent and unrelated items. B) Total time spent fixating on each item of the pair during the 3s encoding trial. More time spent fixating on unrelated C items. * p ≤ .05, ** p ≤ .01, *** p ≤ .001

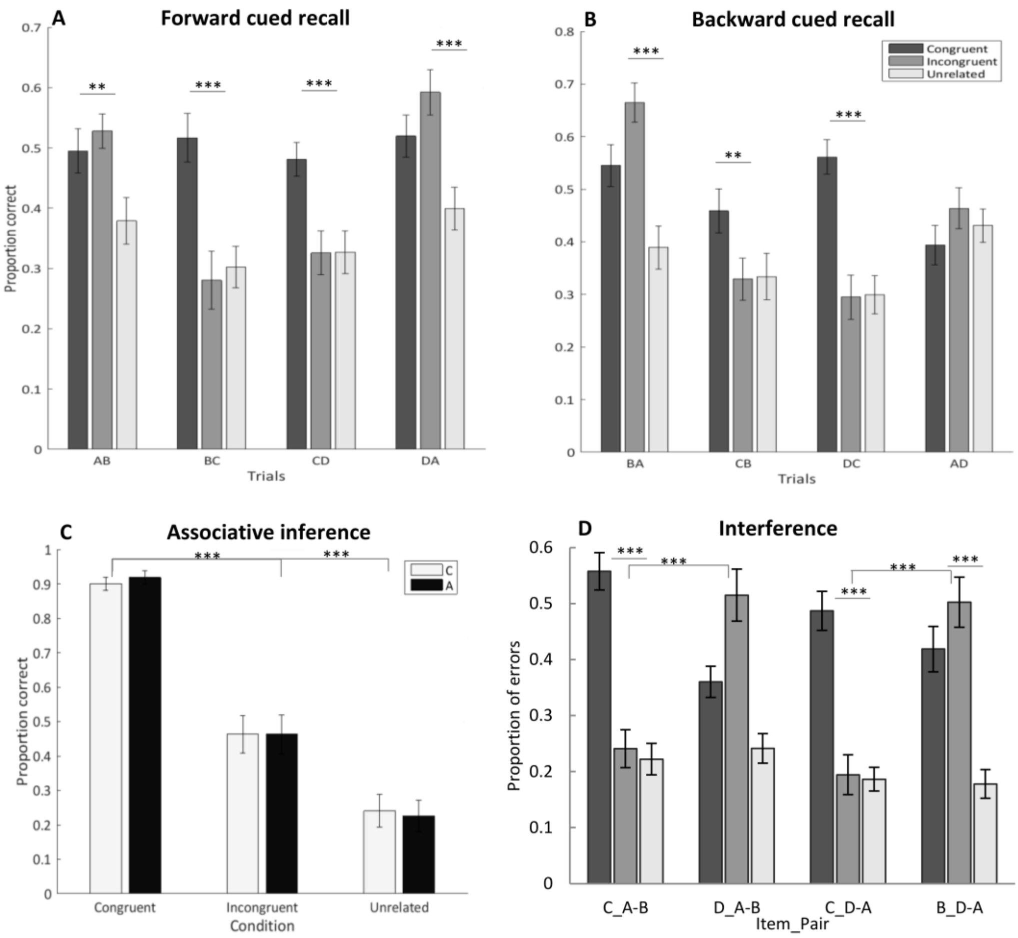

Figure 4 | Behavioural results Experiment 2. A) Cued recall performance forward pairs (e.g. A-?), incongruent items on par with unrelated ones. B) Cued recall performance backward pairs (e.g. ?-B), similarly to forward order, incongruent and unrelated items associated with reduced performance. C) Associative inference, recall of A cued by C and vice versa. Near-ceiling performance for congruent items, followed by incongruent items associated with better performance than unrelated ones D) Interference analysis, higher percentage of erroneously recalled congruent C items in the cued recall task. * p ≤ .05, ** p ≤ .01, *** p ≤ .001