Title
Effect of stimuli congruency on gaze behavior and memory

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 44(44)

Authors
Kraeter, Vivienne
Kritskaia, Veronika
Wiltshire, Travis J.
et al.

Publication Date
2022

Peer reviewed

Effect of stimuli congruency on gaze behavior and memory

Vivienne A. Kraeter (v.a.kraeter@tilburguniversity.edu)
Tilburg University, Department of Cognitive Science & Artificial Intelligence
Warandelaan 2, Tilburg, 5037 AB, the Netherlands

Veronika Kritskaia (v.kritskaia@tilburguniversity.edu)
Tilburg University, Department of Cognitive Science & Artificial Intelligence
Warandelaan 2, Tilburg, 5037 AB, the Netherlands

Travis J. Wiltshire (t.j.wiltshire@tilburguniversity.edu)
Tilburg University, Department of Cognitive Science & Artificial Intelligence
Warandelaan 2, Tilburg, 5037 AB, the Netherlands

Silvy H.P. Collin (s.h.p.collin@tilburguniversity.edu)
Tilburg University, Department of Cognitive Science & Artificial Intelligence
Warandelaan 2, Tilburg, 5037 AB, the Netherlands

Abstract

We investigated whether schema congruency differentially affects low level sensory processing (eye gaze) compared to higher-level cognition (memory). Participants performed a two-phase eye tracking task; first a baseline phase with only congruent cartoon events, and subsequently an experimental phase in which the same events were adapted to remain congruent or become incongruent to a theme. Results revealed that participants became quicker in recognizing the congruent cartoon events compared to incongruent in the experimental phase, indicating improved memory for congruent cartoon events. No mean difference in gaze towards congruent versus incongruent events was observed. Surprisingly, a slight bias towards gazing to the left side of the screen in the baseline phase diminished during the experimental phase, indicating that the schema congruency manipulation might affect gaze behavior. Taken together, our results suggest that our schema congruency manipulation affects gaze behavior and memory, but further eye tracking analysis could reveal the dynamic nature of this effect.

Keywords: memory; eye tracking; event cognition; sensory processing; psychology

Introduction

Organizing structures, such as schemas, are one of the key cognitive mechanisms that allow us to develop knowledge about, and memory of, our environment and the situations we experience. The degree to which these schemas are aligned with reality affects our memory and knowledge formation. Evidence has shown, for example, that information congruent to a schema is remembered better than schema incongruent information (van Kesteren et al., 2012; Ghosh & Gilboa, 2014; Bartlett, 1995). It is important that these schemas are flexible enough to adapt to the ever-changing situations in the world around us. And, indeed, research indicates that our memory has the flexibility to be continuously updated when new information becomes available (Richter et al., 2019; Hupbach et al., 2008; Collin, Milivojevic, & Doeller, 2021). However, besides schema congruent information being better remembered, one could also argue that schema incongruent information, possibly leading to surprise and large prediction error (Reynolds, Zacks & Braver, 2007), could, in some circumstances, improve memory (Frank et al., 2018).

How exactly schema congruency affects memory and knowledge formation could also differ depending on the level at which it is investigated. The model of situation awareness (Endsley, 1995; 2021), which is referred to as perception of objects in the surrounding environment and understanding of their meaning and their role in the future, suggests three levels at which people can have knowledge about what is going on around them: perception, comprehension, and prediction. Could the influence of embedding information in a schema be different across these different levels of situation awareness? To explore this, our research investigates the influence of embedding information in a schema in terms of situation awareness.

How does the first level of situation awareness, perception, relate to schema congruent vs incongruent information? Prior investigations have shown that attention is usually directed to the scene location or region that is semantically most informative and relevant in the particular situation (Võ et al., 2019). Objects in the environment are constrained by certain settings that we have learned in order to understand the surroundings, recognize objects, and interact with them in a goal-directed manner (Võ et al., 2019). More specifically, the same objects tend to appear in similar environmental settings (e.g., books on a bookshelf). People tend to focus their attention longer on inconsistencies within the environment or the objects embedded within it, compared to consistent/congruent settings (Võ et al., 2019). Our study investigates this topic in a particular schematic context, by presenting participants with cartoon images that are either congruent or incongruent with a general schema. Our goal is to identify whether schema congruence influences perception of these cartoon images by analyzing eye gaze patterns and behavior of participants. *We hypothesized that perception will be higher for stimuli that are incongruent with the*

In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society.*

schema compared to congruent, as evident by differences in people's gaze behavior.

How do higher levels of situation awareness relate to schema congruent vs incongruent information? Research suggests that schemas have a profound influence on memory, in most cases, information that is congruent with existing knowledge is remembered better than less congruent information (van Kesteren et al., 2012; Ghosh & Gilboa, 2014). Also, from a neuroscientific perspective, research suggests that the presence of schema is beneficial for memory consolidation (Morris, 2006; Tse et al., 2007). In line with this, *we hypothesized that higher levels of situation awareness are lower for schema incongruent events compared to schema congruent events.* Participants were presented with cartoon images that were either congruent or incongruent with a schema, implemented as a general "theme" (i.e., education). To assess the influence of the schema on memory performance, the study included a memory task consisting of two phases: a baseline phase where the cartoon images were presented before having revealed the schematic context, and an experimental phase after revealing the schematic context (see Figure 1). Both phases consisted of an encoding phase (for baseline without the schematic context and for experimental with the schematic context, see Figure 1) followed by a memory task (on the cartoon images without showing the context for both the baseline phase and the experimental phase). We predicted comprehension of schema-congruent stimuli to be easier than schema-incongruent stimuli, leading to better recognition and faster reaction time for congruent stimuli.

Method

Participants

Seventy Tilburg University students between the ages of 18 and 29 ($M = 22$), composed of 28 men and 42 women, took part in this study. Participants were invited to participate through the SONA participant pool of Tilburg University and were compensated for their participation with course credit (0.5 points for 30 minutes). This study was created and hosted on Gorilla (www.gorilla.sc), an online experiment builder (Anwyl-Irvine et al., 2020). Thus, the study took place remotely, and participants used their own devices and internet connections to participate. The experiment was approved by the local ethical review committee (Tilburg School of Humanities and Digital Sciences, Tilburg University, Netherlands) and participants gave written informed consent.

Given the remote nature, the data were inspected for possible exclusion from the analyses. Participants and individual trials were excluded due to technical difficulties (such as loading errors and eye tracking inaccuracies) along with unusually long trial times due to assumed distraction from the task. Loading errors caused some repeated trials, of which only the first iteration was included for analysis as it was assumed to be the more natural response. Trials from the eye tracking tasks were excluded if accuracy was unsatisfactory as determined using "face_conf" value less than 0.5. The "face_conf" values are based on Support Vector Machine (SVM) classifier scores for the face model fit ranging from 0 (no fit) to 1 (perfect fit). 0.5 or higher are acceptable values (Gorilla, n.d). If more than 25% of trials of a participant's eye tracking accuracy were unsatisfactory, the entire participant was excluded from further analysis in either phase as results in the memory tasks are dependent on the successful completion of the eye tracking tasks. Furthermore, participants were excluded from both phases if there was missing data in the eye tracking tasks. Sixteen participants were excluded based on these criteria, and therefore the subsequent analyses contained data of 54 participants. Additionally, trials in the memory tasks with unusually lengthy response times (over 60s) were excluded from analysis due to the presumption that factors beyond the experiment influenced the participant's time of response.

Materials

Two sets of stimuli were presented to participants during this study. The first set (presented in the baseline phase) contained images of a cartoon character performing actions with props which could typically be found in a classroom setting. The character therefore could appear to be a teacher, which could be perceived as congruent to the general episodic theme of education. These images contained no background, and thus it was not clear in the baseline phase whether the original events were to be incongruent (Figure 1.A) or congruent (Figure 1.C) to the theme. In the experimental phase, the original events were updated with schematic context in the form of an added background to create the second set of stimuli. These modified events revealed whether the cartoon character was in fact a teacher in a classroom setting, and therefore whether an image from the first set was indeed incongruent (Figure 1.B) or congruent (Figure 1.D) to the general episodic theme of education. To account for potential gaze biases towards humans and faces, character location and quantity were counterbalanced.

Procedure

The experiment followed a within-subject experimental design in which the differences in perception and situation - awareness when stimuli were congruent or incongruent to a general episodic theme was explored. Participants first provided consent, before completing a basic demographic questionnaire.

The first (baseline) phase started with an encoding task, in which participants were shown two images at a time and asked to memorize the contents. Prior to each encoding task, the webcam-based eye tracking was calibrated. No information about the general 'education' theme was provided to participants. Images were pseudo-randomized to comprise three pair types: i) trials with two congruent images (CC), ii) mixed trials in which a congruent image was on the left and an incongruent image was on the right (CI), and iii) mixed trials in which an incongruent image was on the left and a congruent image on the right (IC). Participants were shown 24 side-by-side sets of images (10 CC pairs, 7 CI pairs, and 7 IC pairs) for 5s each, with a fixation cross in the center of the screen separating each stimuli pair for 500ms and 100ms of white screen only for 100ms before and after the cross. These baseline images depicted the original events, before the context, and therefore could all be considered congruent with the general episodic theme of education as their context was not yet revealed. While looking at the stimuli, participants' eye gaze coordinates were estimated through their webcam using Gorilla's eye tracking capabilities with maximum sampling rates dependent on the participants' monitor refresh rate.

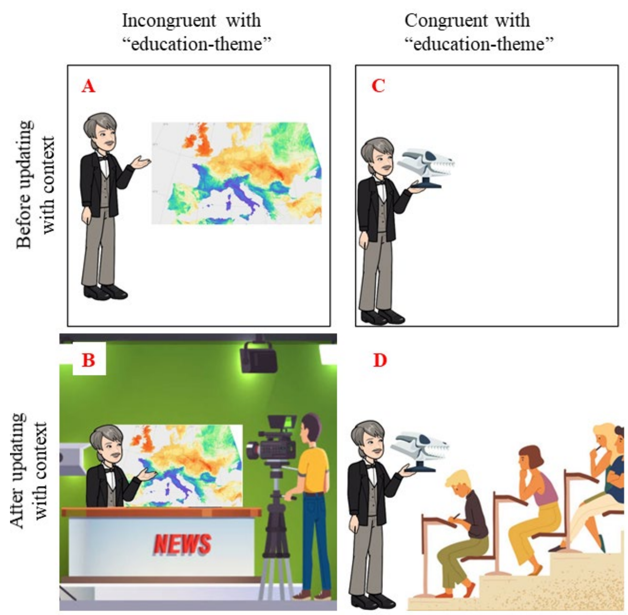

Figure 1: Stimulus examples. A: event before adding context that will later be revealed as incongruent with the theme. B: event after adding context which is incongruent with the theme. C: event before adding context that will remain congruent with the theme. D: event after adding context that is congruent with the theme.

Subsequently, participants were asked to perform a 10-item math test as a distractor task. Participants had to solve basic mathematical problems and insert the correct answer within 5s. This distractor task provided a break between the encoding and memory tasks, ensuring that rehearsal of the presented stimuli was limited before participants were asked to recall them.

After the encoding and distractor tasks, participants completed a memory task in which they were shown one image at a time and prompted to recall whether the presented image is new or old. Images were randomly shuffled, and displayed in the center of the screen. Along with determining whether images were presented earlier (during encoding), participants were asked to rate their confidence in their answer (1: very low – 4: very high). The memory task was self-paced, and participants' reaction times to their answers and confidence ratings for 48 images (24 congruent, of which 7 would be revealed as incongruent in the experimental phase, and 24 new) were recorded.

The second (experimental) phase consisted of the same tasks, with the main difference being the stimuli shown during encoding. This time, participants were shown the modified events, updated with context schema through added backgrounds. As in the baseline phase, images were pseudo-randomized and presented for 5s in the aforementioned pair types: (CC), (CI), and (IC). The added context schema allowed for updating of the stimuli encoded in the baseline phase, as the congruence (or incongruence) to the general episodic theme of education was revealed.

As in the baseline phase, encoding was followed by another 10-item timed math test distractor task and a subsequent memory task. In the experimental memory task, participants were presented with the same stimuli as in the baseline phase (images with no schematic context), which were randomly shuffled to differ from the order presented in the baseline phase. Performing the memory tasks of both phases on the same set of stimuli allowed for testing of the underlying schema in the participants' memory, which they became aware of during encoding in the experimental phase. Therefore, the first phase provided baseline information, while the experimental phase showed how the underlying schema and knowledge of images being congruent and incongruent to the episodic theme impacted participants' eye gaze behavior, level of recognition, and reaction times.

Statistical Analyses

In order to test the hypotheses of this study, three metrics were created: average eye gaze percentage per side of the screen, level of recognition, and reaction time during the memory task. Participants' eye gaze behavior (how long they looked at congruent and incongruent stimuli on the left or right of the screen) during encoding was used to assess the hypothesis that perception is higher for incongruent than for congruent events. This metric was used to determine whether the underlying schema led participants to look more at incongruent than congruent stimuli during encoding. The level of recognition (the percentage of correctly classified images as new or old) and reaction times during the memory task were used to assess the second hypothesis, that higher levels of situation awareness are lower for incongruent than congruent events. This allowed for testing if congruent stimuli were recalled and recognized at a higher rate and speed than incongruent stimuli.

Each metric was baseline-corrected by subtracting the values of the experimental phase by the values of the baseline phase, creating difference scores. Skewness, kurtosis, and Shapiro-Wilk tests were conducted for all metrics' difference scores to determine whether they followed a Gaussian distribution. As all metrics were non-Gaussian, the difference scores for congruent and incongruent stimuli of each metric were compared using a Wilcoxon signed-rank test. These analyses examined whether underlying schema led to differences between congruent and incongruent stimuli for each metric. Code and data for all analyses is available here: https://osf.io/2xg8f/.

Results

To assess participants' higher levels of situation awareness, reaction time to participants' answers in the memory task were examined. In the experimental phase (2), when correcting for the baseline phase (1), participants responded faster to congruent (M = -549.44, SD = 610.95) compared to incongruent (M = -318.20, SD = 776.08) events (t = 493, p = 0.01, r_{rb} = -0.38, 95% CI [47.02, 362.58])[1]. This effect is illustrated in Figure 2, and an overview of mean reaction times, prior to baseline-correction, can be found in Table 1.

This difference between congruent and incongruent stimuli in the memory task was not found for level of recognition (congruent: M = 0.84, SD = 8.12; incongruent: M = 1.53, SD = 17.59; t = 468, p = 0.58, r_{rb} = -0.10, 95% CI [-2.94, 4.20]). As seen in Table 2, a ceiling effect was observed as level of recognition was generally high across both phases.

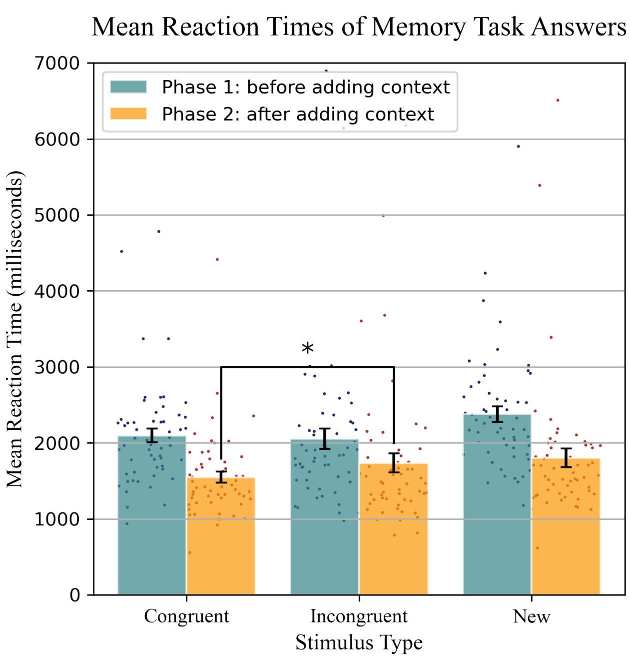

Figure 2: Comparing mean reaction times to memory task answers to congruent, incongruent, and new stimuli, for the baseline phase (1) and experimental phase (2), respectively. Note that stimuli labeled as incongruent in phase 1 had not yet been revealed as such, and were therefore perceived as congruent to the theme until their context was revealed in phase 2.

[1] Levels of significance vary slightly depending on participant inclusions/exclusions based on the exclusion rules specified in the methods section. The general pattern of results, however, is robust.

Table 1: Mean reaction times (ms) in the memory tasks.

Stimulus type		Reaction Time (ms)	
		Phase 1	Phase 2
Congruent	Mean	2096.95ms	1547.51ms
	Median	2007.35ms	1424.77ms
Incongruent	Mean	2052.62ms	1734.42ms
	Median	1842.37ms	1480.19ms
New	Mean	2378.81ms	1802.52ms
	Median	2223.73ms	1593.19ms

Table 2: Levels of stimulus recognition in memory tasks.

Stimulus type		Level of Recognition	
		Phase 1	Phase 2
Congruent	Mean	93.17%	94.01%
	Median	94.12%	100%
Incongruent	Mean	89.03%	90.56%
	Median	100%	100%
New	Mean	87.96%	95.31%
	Median	89.58%	100%

To assess low level of situation awareness, perception, of congruent and incongruent stimuli, the participants' gaze behavior during encoding was analyzed to determine whether the introduction of stimuli incongruent to the general episodic theme of education influenced gaze behavior. No difference between the percentage of time looked at incongruent stimuli (M = 0.93, SD = 24.27) versus congruent stimuli (M = -0.93, SD = 24.27) in mixed trials was found (t = 3014.50, p = 0.66, r_{rb} = -0.05, 95% CI [-6.71, 9.81]). Table 3 presents mean and median percentages of time looked at each stimulus type, and whether the stimulus was on the left or right side of the screen.

When examining participants' gaze behavior further, a general difference was observed in the percentage of time looked at the left versus the right side of the screen when comparing the baseline to the experimental phase. A slight bias towards looking at the left side of the screen was present in the baseline phase (i.e., before adding context) across all trial types (CC trials, mixed trials congruent-incongruent, and mixed trials incongruent-congruent) (M = 53.64%), which decreased in the experimental phase (M = 51.07%) (i.e., after context was introduced). This effect is illustrated in Figure 3.

Table 3: Eye gaze behavior.

Stimulus type	Phase		Left %	Right %
Values for stimuli of **control trials** (congruent-congruent):				
CC Pairs	**Phase 1**	Mean Median	54.20% 53%	45.80% 47%
	Phase 2	Mean Median	52.53% 54%	47.47% 46%
Values for stimuli of **mixed trials** (congruent-incongruent and incongruent-congruent):				
Congruent	**Phase 1**	Mean Median	55.10% 54%	48.38% 49%
	Phase 2	Mean Median	51.12% 51%	50.43% 52%
Incongruent	**Phase 1**	Mean Median	51.62% 51%	44.91% 46%
	Phase 2	Mean Median	49.57% 48%	48.88% 49%

Note: percentages may add up to over 100% due to rounding.

Discussion

This study examined how the presence of an underlying schema can differentially affect gaze behavior and episodic memory through a two-phase, within-subjects experiment. Participants' memory for semi-naturalistic cartoon images was tested before and after the introduction of an underlying schematic context while their gaze behavior was recorded. Both gaze behavior during encoding as well as participants' subsequent memory performance was analyzed to assess the influence of an underlying schema on low and high levels of situation awareness (Endsley, 1995). Results revealed that, despite no significant difference in recognition between schema-congruent and schema-incongruent cartoon images, participants recognized schema-congruent cartoon images faster than incongruent cartoon images. Furthermore, we did not observe a mean difference in eye gaze to congruent vs incongruent cartoon images during encoding. However, surprisingly, we did observe a bias to gazing to the left side of the screen during the baseline phase, which diminished once the congruent/incongruent context had been revealed.

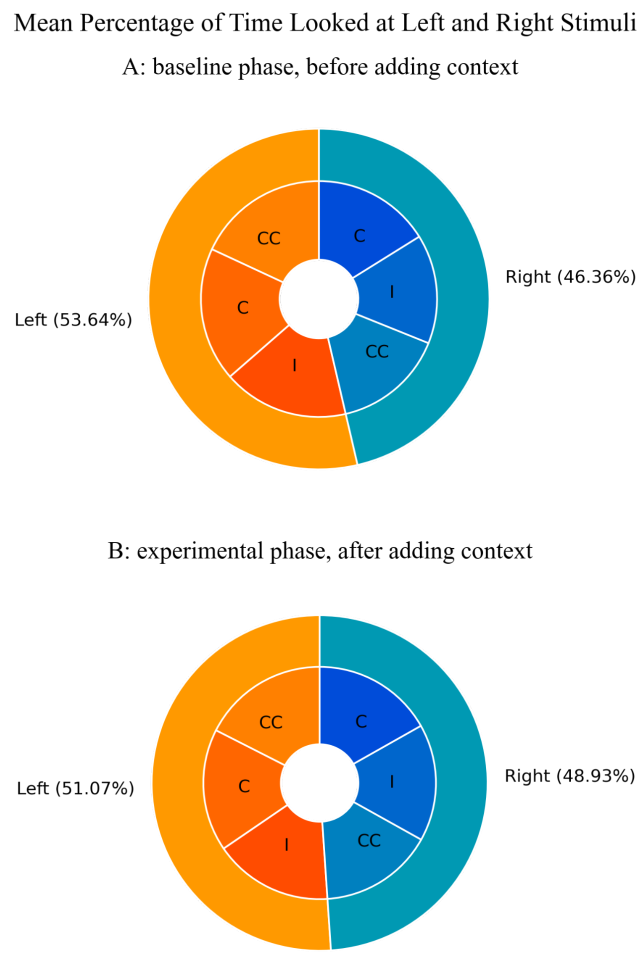

Figure 3: Participants' gaze behavior before and after schematic context is revealed.

Participants' general level of recognition and reaction time in the memory task suggests that participants needed longer to comprehend the cartoon images incongruent to the underlying schematic context. *This confirms our hypothesis that higher levels of situation awareness (e.g., comprehension) is lower for incongruent cartoon images compared to congruent cartoon images*. This finding is in line with earlier research that showed worse recognition and recall results for incongruent events compared to congruent events (Reynolds, Zacks & Braver, 2007; Frank et al., 2018), which was supported by eye tracking analysis of fixation patterns during encoding of congruent and incongruent stimuli (Frank et al., 2018). Furthermore, Frank et al. revealed increased fixation on first unexpected incongruent stimuli, suggesting that it is more difficult to comprehend and encode events that are unexpected and incongruent with the underlying schema (Frank et al., 2018). This suggests that further statistical analysis of our eye tracking data will give insight on how people analyze visual stimuli, taking into account congruency of the schema, and evaluate if there are changes in gaze patterns that would suggest an updating of the schema.

As mentioned before, in the experimental phase participants neither had a preference for a certain side of the screen nor did they have a preference for a particular type of image, when looking at mean values of eye gaze, suggesting low levels of situation awareness (i.e., perception) were not influenced by the schema. However, before revealing the underlying schematic context, in the baseline phase, there was a slight bias to gaze at the left side of the screen, suggesting there might have been an influence of schematic context on gaze behavior. Such preference for the left side of the screen, in the baseline phase, can be justified by the way we read, from left to right, suggesting that participants tend to look to the left slightly more often than to the right and start to browse the screen in the same way as reading. Surprisingly, after revealing which stimuli were congruent and which stimuli were incongruent to the schematic context, this slight bias decreased. More fine-grained measures could help to understand this possible influence of schematic context on eye gaze differences for incongruent vs congruent cartoon events. Further work could aim to identify meaningful changes in the dynamics of the eye gaze, for example, by adopting methods used by Stephens et al (2018). In particular they were trying to capture real-time dynamics of cognition using eye-tracking data analyses, like changes in entropy and power-law behavior. They found that changes in entropy and changes in power-law behavior predict a phase transition in the cognitive systems (Stephens et al., 2009). This is a promising direction to identify behavioral patterns that can be connected with broader theory concepts about the effects of schema on gaze behavior and memory updating. In our case, analyses like this might give insight on how people analyze/scan the visual stimuli, also taking into account the congruency of the image and evaluating if there are any key moments of change, which could suggest a potential updating of the schema. Furthermore, an analysis of, for example, fixation patterns when schema-congruent and schema-incongruent images are presented would identify if there are any fixation points that could be affected by the presence of schema (Frank et al., 2018).

It is important to acknowledge some common limitations of online studies, considering the lack of supervision of participants and lack of control of outside factors that could have affected the dataset. However, this does not undermine the quality of the data, as previously researched, online data collection provides reliable results that meet or exceed the standards of published research (Buhrmester et al., 2011).

In conclusion, this study provides further evidence for the assumption that there is a relationship between episodic memory updating, schema, and different levels of situation awareness. Evidence from eye tracking and reaction time data, in particular, suggests that this relationship should be explored further to better understand the dynamic interplay of schemas, episodic updating, and behavior.

Acknowledgements

The authors thank Neil Cohn for a helpful discussion on the design of this experiment.

Author Contribution Statement

Vivienne A. Kraeter: Software, Writing - Original Draft, Visualization, Formal Analysis, Data Curation. **Veronika Kritskaia:** Data Curation, Writing - Original Draft. **Travis J. Wiltshire:** Conceptualization, Methodology, Writing - Reviewing & Editing, Supervision, Resources, Investigation. **Silvy H.P. Collin:** Conceptualization, Methodology, Resources, Writing - Reviewing & Editing, Supervision, Resources, Investigation, Visualization.

References

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods, 52*(1), 388–407.

Bartlett, F. C. (1995). Remembering: A study in experimental and social psychology. Cambridge University Press.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data? *Perspectives on Psychological Science*, 6(1), 3–5.

Collin, S. H. P., Milivojevic, B., & Doeller, C. F. (2021). Hippocampal reconfiguration of events in mnemonic networks. *BioRxiv*, 2021.05.25.445607.

Endsley, M. R. (1995). Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors, 37*(1), 32–64.

Endsley, M. R. (2021). A Systematic Review and Meta-Analysis of Direct Objective Measures of Situation Awareness: A Comparison of SAGAT and SPAM. *Human Factors*, 63(1), 124–150.

Frank, D., Montaldi, D., Wittmann, B., & Talmi, D. (2018). Beneficial and detrimental effects of schema incongruence on memory for contextual events. *Learning & Memory*, *25*(8), 352-360.

Ghosh, V. E., & Gilboa, A. (2014). What is a memory schema? A historical perspective on current neuroscience literature. *Neuropsychologia*, *53*, 104-114.

Gorilla. (n.d.). Eye Tracking in Gorilla. Retrieved from Gorilla: https://support.gorilla.sc/support/reference/eye-tracking#analysis

Hupbach, A., Hardt, O., Gomez, R., & Nadel, L. (2008). The dynamics of memory: Context-dependent updating. *Learning & Memory*, *15*(8), 574-579.

Morris, R.G. (2006) Elements of a neurobiological theory of hippocampal function: the role of synaptic plasticity, synaptic tagging and schemas. *European Journal of. Neuroscience, 23,* 2829–2846

Reynolds, J. R., Zacks, J. M., & Braver, T. S. (2007). A computational model of event segmentation from perceptual prediction. *Cognitive science*, *31*(4), 613-643.

Richter, F. R., Bays, P. M., Jeyarathnarajah, P., & Simons, J. S. (2019). Flexible updating of dynamic knowledge structures. *Scientific Reports*, *9*(1), 1-15.
Shiferaw, B., Downey, L., & Crewther, D. (2019). A review of gaze entropy as a measure of visual scanning efficiency. *Neuroscience & Biobehavioral Reviews*, *96*, 353-366.
Stephen, D. G., Boncoddo, R. A., Magnuson, J. S., & Dixon, J. A. (2009). The dynamics of insight: Mathematical discovery as a phase transition. *Memory & Cognition*, *37*(8), 1132-1149.
Tse, D., Langston, R. F., Kakeyama, M., Bethus, I., Spooner, P. A., Wood, E. R., Witter, M. P., & Morris, R. G. M. (2007). Schemas and memory consolidation. *Science*, *316*(5821), 76–82.
Van Kesteren, M. T., Ruiter, D. J., Fernández, G., & Henson, R. N. (2012). How schema and novelty augment memory formation. *Trends in Neurosciences*, *35*(4), 211-219.
Võ, M. L. H., Boettcher, S. E., & Draschkow, D. (2019). Reading scenes: how scene grammar guides attention and aids perception in real-world environments. *Current Opinion in Psychology*, *29*, 205-210.