



Emotional enhancement and asymmetric context congruency effects for incidental memories of 2D and VR scenes

Marlene Wessels¹ · Christian Valuch² · Anne Schacht² · Lousia Kulke³

Received: 10 September 2024 / Accepted: 26 January 2026
© The Author(s) 2026

Abstract

Virtual reality (VR) has become more affordable and is increasingly used for therapeutic and educational purposes where it is important to transfer learned experiences to the real world. But is it worth switching from conventional 2D displays to VR to strengthen such learning effects? Here, we investigated how well incidentally learned emotional and neutral scenes, encoded either in VR or as a conventional 2D presentation (encoding context VR, 2D), could be recognized in congruent or incongruent presentation contexts (retrieval context VR, 2D). Using a $3 \times 2 \times 2$ mixed design, 60 participants viewed (and incidentally encoded) emotional and neutral scenes in VR and as 2D video. After 24 h, 30 participants performed an unannounced old-new judgment task in 2D, while the others completed it in VR. 2D-encoded scenes were recognized better in the congruent 2D than in the incongruent VR context. Context congruency effects were less pronounced for VR-encoding. On average, participants reliably recognized VR-encoded scenes in the incongruent 2D retrieval context. Participants retrieved emotional scenes better and more confidently than neutral ones, independent of encoding context. However, scenes experienced in VR were more likely to be perceived as emotional than when the same scenes were viewed as 2D video. Taken together, the presentation mode of a scene (2D, VR) served as a contextual mnemonic aid, mainly for 2D-encoded scenes. Overall, VR-encoded scenes were less dependent on context congruency effects, suggesting a reduced sensitivity to the retrieval context rather than a direct transfer advantage. Finally, VR-encoding increased the likelihood that the content would be perceived as emotional, suggesting its potential as a useful educational and therapeutic tool.

Keywords Virtual reality · Incidental memory · Context dependence · Recognition · Emotion

1 Introduction

In recent years, the field of therapeutic and educational interactions has undergone a dual transformation. First, the increased affordability of VR technology has significantly elevated its use in these settings. Concurrently, the COVID-19 pandemic necessitated a shift towards online interactions (Lemay et al. 2021), where content is typically shown on a conventional 2D display. These two modes of

presentation—2D and VR—offer varying levels of immersion (i.e., objective degree of realism of a simulation), with VR potentially blurring the boundaries between virtual and real experiences (Bonnail et al. 2024; Rubo et al. 2021). Given that the effectiveness of learning or therapeutic interventions likely relies on the successful retrieval of acquired knowledge within real-world contexts (Clay et al. 2024), differences in immersion between 2D and VR presentations may influence intervention effectiveness. Memory retrieval is enhanced when encoding and retrieval contexts align, whereas incongruence increases transfer costs—that is, the additional cognitive effort or performance decline occurring when retrieval and encoding contexts do not match, as described by the encoding specificity principle and transfer-appropriate processing in numerous studies (Tulving and Thomson 1973; Godden and Baddeley 1980; Watkins et al. 1976; Shin et al. 2021; Choi et al. 2025; S. M. Smith and Vela 2001; Godden and Baddeley 1975). While some studies have revealed inconsistencies in replicating such effects

✉ Marlene Wessels
mwessels@uni-mainz.de

¹ Institute of Psychology, Johannes Gutenberg-University Mainz, Wallstraße 3, 55122 Mainz, Germany

² Institute of Psychology, Georg-August-University of Göttingen, Göttingen, Germany

³ Department of Developmental Psychology With Educational Psychology, University of Bremen, Bremen, Germany

within real-world and VR contexts (Murre 2021; Wälti et al. 2019), other research showed that switching between VR and real-world environments or between two-dimensional and three-dimensional VR contexts negatively affected memory recall (Johnsdorf et al. 2025; Lamers and Lanen 2021). This suggests that the presentation mode itself may serve as a contextual cue during retrieval. Thus, highly immersive VR experiences that are more similar to real-world experiences could potentially be better retrieved in the real world than observed 2D presentations (Martarelli et al. 2023). This hypothesis prompts an inquiry into whether the mode of presentation, 2D screen or VR, significantly impacts memory encoding and retrieval, a question that is central to our study.

Previous research has indicated differences in memory processes between VR- and 2D-encoded scenes (Allcoat and Mühlenen 2018; Johnsdorf et al. 2023a; Kisker et al. 2020; Wälti et al. 2019). Some studies indicate improved memory performance for various types of intentionally learned content presented in VR as compared to 2D screens. The learning content comprised, for example, navigation routes (Kim et al. 2018), objects (Mania et al. 2003), or faces (Krokos et al. 2019). This benefit of VR-based learning may be explained by a more active exploration of the virtual environment or higher immersion (S. A. Smith 2019; van Helvoort et al. 2020). Even when participants were not explicitly instructed to memorize the learning content and formed so-called incidental memories, which are more generalizable to real-world situations, virtual experiences were associated with a mnemonic advantage over 2D presentations. This was demonstrated, for example, in a study by Schöne and colleagues (2019a), in which participants recognized snapshots of a previously encoded VR motorcycle ride more accurately than those of a 2D ride. In addition, this advantage was demonstrated in a recognition task in which participants were presented with specific objects that they had previously seen and thus incidentally learned in either a VR or 2D environment (Kisker et al. 2019b). Although both encoding groups performed similarly on average, the VR-encoding group recognized the objects more vividly. In contrast, the 2D group reported only a sense of familiarity during retrieval. In sum, VR experiences are assumed to be linked to more personally relevant autobiographical memories, whereas 2D-encoded scenes presumably form shallower episodic memories that are less connected to one's sense of self (Conway 2005; Roediger and Marsh 2003).

However, some previous studies (Kisker et al. 2019b; Schöne et al. 2019a) may not have captured the full quality of VR memories, as recognition tasks were conducted exclusively in a 2D format, potentially underestimating memory fidelity. Conversely, the consolidation of 2D-encoded presentations might have been overestimated due to congruent

2D retrieval contexts. Our study addresses this issue by additionally incorporating a VR recognition task, allowing us to thoroughly investigate whether the presentation mode (2D screen, VR) is a critical contextual factor in facilitating memory retrieval.

Even if a congruent presentation mode of encoding and retrieval contexts does not enhance the recall of memory, incongruent retrieval conditions could still cause transfer costs. VR-encoded scenes presumably create autobiographical memories (Gupta et al. 2022; Kisker et al. 2019b, 2020; Schöne et al. 2019a), which are linked to the individual self and are organized in an extensive network of different abstractions (Conway 2001). This depth of encoding prompts that they may be less susceptible to transfer costs in a retrieval context that diverges from the original VR-encoding context, in contrast to more superficial memories associated with observed 2D screen presentations. Alternatively, while the transfer costs when retrieving VR-encoded and 2D-encoded scenes might prove similar in terms of memory performance, the confidence in recognizing previously encountered stimuli could still be significantly higher within a VR-encoding context. This notion aligns well with the idea of a more vivid and effortless recollection associated with (autobiographical) VR memories (Kisker et al. 2019b, 2020).

Importantly, VR experiences also tend to elicit lifelike emotions (Chirico and Gaggioli 2019; Gupta et al. 2025; Kisker et al. 2019a; Parsons 2015; Somarathna et al. 2023), increasing ecological validity and potentially surpassing the emotional impact of 2D presentations (Estupiñán et al. 2014). Emotions are typically characterized along two key dimensions: valence and arousal. Valence refers to the positivity or negativity of an emotional experience, whereas arousal reflects the intensity or activation level of emotion (e.g., calm vs. excited) (Adelman and Estes 2013; Bradley and Lang 1994). Given that emotional content, whether negative or positive, is known to generate more elaborate memory traces than neutral content (Wardell and Palombo 2024; Williams et al. 2022; Gao et al. 2024; Dolcos and Cabeza 2002; Harris and Pashler 2005; Adelman and Estes 2013), the emotional richness of VR could further facilitate successful memory retrieval (Gupta et al. 2025). This phenomenon is referred to as the emotional enhancement effect, which means that emotional experiences are often remembered better than neutral ones (Chainay et al. 2012; Hamann 2001). In fact, the study by Schöne and colleagues (2019a) revealed a higher positive correlation between the mood experienced after the motorcycle ride and the performance in the recognition task in the VR-encoding group than in the 2D-encoding group. This suggests that the same scenario might have evoked more emotionally profound experiences in VR than in 2D, which could have led to the

successful retrieval of VR experiences. Could the emotional enhancement effect on memory be more pronounced for VR-encoded scenes than for 2D events? Two studies examined the emotional enhancement effect on memory for objects encoded in VR (Cadet et al. 2021; Cadet and Chainay 2020). They found better performance for the immediate recall of emotionally connotated objects previously presented in VR, compared to neutral objects. However, the emotional valence was determined by a separate pre-test sample in these studies. Therefore, the emotional enhancement effect for VR and 2D scenes still remains to be investigated by considering the actual emotional responses of the participants who performed the memory task.

In sum, previous studies have primarily focused on the differences in encoding contexts by comparing 2D and VR presentations, and did not assess participants' actual emotional responses during encoding. Moreover, the interaction between emotional enhancement effects and retrieval context has not yet been systematically examined across different presentation modes. Our study addresses this gap by providing a more comprehensive understanding of how incidentally learned, emotional content, as indexed by subjective valence and arousal ratings during encoding, influences subsequent recognition performance and confidence in encoding-congruent and -incongruent contexts. Gaining insight into the interplay between encoding modality, emotional experience, and retrieval context could guide the design of immersive technologies for more effective learning and therapeutic interventions (Kredlow et al. 2018). For instance, in VR exposure therapy, which is used to treat anxiety disorders and post-traumatic stress disorder, patients are gradually exposed to fear-inducing stimuli in immersive virtual environments, with the goal of transferring learned coping strategies to the real world (Carl et al. 2019; Diemer et al. 2014; Gonçalves et al. 2012; Kothgassner et al. 2019; Schröder et al. 2023; Shahid et al. 2024; Wong et al. 2023; Wu et al. 2021). Similarly, VR-based emotion regulation training places individuals in emotionally evocative yet controlled scenarios to help them practice reappraisal or emotional distancing techniques (Montana et al. 2020; Roa and Rodríguez 2024; Yu et al. 2023). The success of such interventions depends on the reliable retrieval of insights acquired during VR sessions in contexts that differ from the original encoding context.

Therefore, we designed the present study with two objectives: 1) the investigation of transfer costs for VR- and 2D screen-encoded scenes (as indicated by a lower context congruency effect in memory), and 2) the valid investigation of the emotional enhancement effect in VR and 2D presentation contexts. The study consisted of an encoding phase, in which each participant rated both VR experiences and 2D scenes in terms of emotional valence and arousal, and

an unannounced old-new recognition task and confidence rating on the following day. Half of the participants completed the recognition task on a 2D screen, while the other half completed it in VR. Thus, retrieval context was introduced as the only between-subjects factor. We compared the recognition performance and the recognition confidence for each emotional category, encoding context (VR, 2D screen) and retrieval context (congruent, incongruent with encoding). We assumed that memory would generally be better if encoding and retrieval contexts for a scene remain congruent compared to situations where a transfer between contexts is necessary (Hypothesis 1). Importantly, we hypothesized that VR experiences form more robust memory traces and should thus produce less pronounced transfer effects compared to scenes that were encoded in 2D but are later revisited in VR (Hypothesis 2). Finally, we expected a more intense emotional response to VR content compared to 2D content and hypothesized that this creates more robust memory for emotional VR compared to emotional 2D scenes, which should be reflected in an interaction between encoding presentation mode and emotional content (Hypothesis 3). In general, we expected emotional content to be better recognized than neutral content.

2 Methods

2.1 Participants

Based on the effect of the encoding presentation mode reported in a previous study (Schöne et al. 2019b), an a priori power analysis was conducted. We conducted separate calculations for each retrieval condition (2D and VR) under the assumption of a within-subjects design to ensure sufficient power to detect the VR superiority effect within each retrieval group, which suggested a required sample size of $n=23$ per group ($d=0.713$, $\alpha=0.05$, $Power(1-\beta)=0.95$). To avoid overestimating the reported effect size, the planned sample size was set to 30 in each of the two retrieval groups. All 60 participants (2D retrieval group: $n=30$; age: $M=23.80$ years, $SD=1.85$ years, 12 women, 18 men, 27 right-handed; VR retrieval group: $n=30$; age: $M=24.60$ years, $SD=2.65$ years, 17 women, 13 men, 28 right-handed) were healthy non-psychology students with normal or corrected-to-normal vision. In case participants needed a visual aid, they wore contact lenses to ensure the optimal fit of the VR goggles. None of them reported a history of neurological or psychological disorders. Exclusion criteria were a current influence of strong fatigue, medication, alcohol or other drugs that could possibly interfere with cognitive processes. Since the use of VR technology is considered to sometimes elicit motion sickness (Parsons

2015), participants who reported a vulnerability for motion sickness (e.g., regular experience of motion sickness while traveling) were excluded from participation. All participants received a monetary reward of €12 (for 1.5 h).

The study was approved by the local ethics committee of the Georg-August-University Göttingen (file number 238), complied with the standards of the Declaration of Helsinki (1964), and was preregistered with the Open Science Framework (2D retrieval group: Wessels et al. 2019a: osf.io/fujpr; VR retrieval group: Wessels et al. 2019b: osf.io/z76sj).

2.2 Material and apparatus

The experiment consisted of two parts, an encoding session and a retrieval session, which took place in two different rooms. In the encoding session, 2D and VR dynamic panoramic scenes were presented. 2D scenes were large monoscopic video projections on a white wall (video size: 88 cm × 173 cm, visual angle: 61.64° × 25.87°). VR scenes were stereoscopic immersive presentations via a head-mounted display (Oculus Rift Developer Kit 2; resolution: 1920 × 1080 pixels, 960 × 1080 pixels per eye, field of view: 100°). The tracking system of the head-mounted display enabled the participants to fully experience the presented virtual scenes through 360° head movements. The encoding session, including the presentation of 2D and VR scenes, was implemented with Unity (version: 2018.2.15). The presented audiovisual scenes ($s=48$, see also Appendix Table A1) of the encoding session were selected from the Library for Universal Virtual Reality Experiments (luVRe) (Schöne et al. 2021). The luVRe database includes various 4K scenes, standardized to a duration of 30 s, which can be projected either as a stereoscopic scene in VR or as a 2D scene. Auditory signals were played using two Samson MediaOne 4a Active Studio loudspeakers. The scenes were selected in a pilot study and aimed to balance the three valence categories across the encoding contexts. The procedure is described in detail in the report on the pilot study (https://osf.io/827jt/?view_only=cc2511ec36ff468fb2c9d8ec74248643).

In the retrieval session, the 2D recognition task (2D retrieval group) comprised the presentation of static panoramic (360°) images ($s=96$) on a laptop screen (image size: 1280 × 600 pixels, screen size: 28.5 × 17.9 cm, screen resolution: 2560 × 1600 pixels, approx. viewing distance: 50 cm) and was implemented using PsychoPy (version: 1.85.4). Half of the static images were target stimuli depicting a static frame from a previously presented scene (“old”). The other half were distractor stimuli depicting a static frame of a scene, which was not presented in the encoding session and was hence “new” to the participants. Distractor stimuli were also derived from the luVRe database. Distractor and

target stimuli used in the recognition task were presented in random order and were matched for visual similarity across the three *Predetermined Valence* conditions in a pilot study (see also https://osf.io/827jt/?view_only=cc2511ec36ff468fb2c9d8ec74248643). For the VR recognition task (VR retrieval group), the same old-new stimuli were presented in VR, and participants could fully explore them by moving their heads. In both retrieval conditions, the content was 360° in nature; however, only the VR retrieval group allowed immersive viewing through head movements.

Note that in VR, participants freely explored the scene during the encoding phase. The perception of the scene depended on the head movements, which were likely to differ between encoding and retrieval phases, even for the same participant. For 2D-encoding and retrieval, we introduced variation in the scene display by switching from a large video projection to a smaller computer screen for the two sessions (Fig. 1). This was done to approximate the inherent variability of visual input in the VR condition, where participants freely explored the scenes through head movements and thus experienced slightly different viewpoints during encoding and retrieval.

2.3 Experimental design

The study comprised a fully-crossed 3 (subjective valence) × 2 (encoding context) × 2 (retrieval context) mixed design. Each participant assessed the subjective valence (negative, neutral, positive) for each scene during the encoding phase and experienced each encoding context (VR, 2D). While the order of the encoding scenes was randomized per participant, the order of the two encoding presentation modes of the scenes was counterbalanced across participants in a block design; one half of the participants started with the VR presentation, the other half started with the 2D presentation. The retrieval context (congruent, incongruent to encoding context) was implemented as a between-subjects factor with participants being evenly and randomly assigned to the two retrieval groups to complete the recognition task in the respective context. To clarify, 2D refers to monoscopic presentation (encoding: video projection, retrieval: laptop screen), and VR refers to stereoscopic immersive presentation via a head-mounted display.

In a previous first pilot study, a different sample ($n=30$, age: $M=23.9$ years, $SD=2.75$ years, 15 women, 15 men, 27 right-handed) rated all scenes in terms of valence and arousal using a digital version of the 5-point Self-Assessment Manikins (SAM) (Bradley and Lang 1994). The pilot study provided a first impression of the approximate valence (negative, neutral, positive *Predetermined Valence*) of the scenes and thus allowed for an approximate counterbalancing of the valence conditions across the presentation

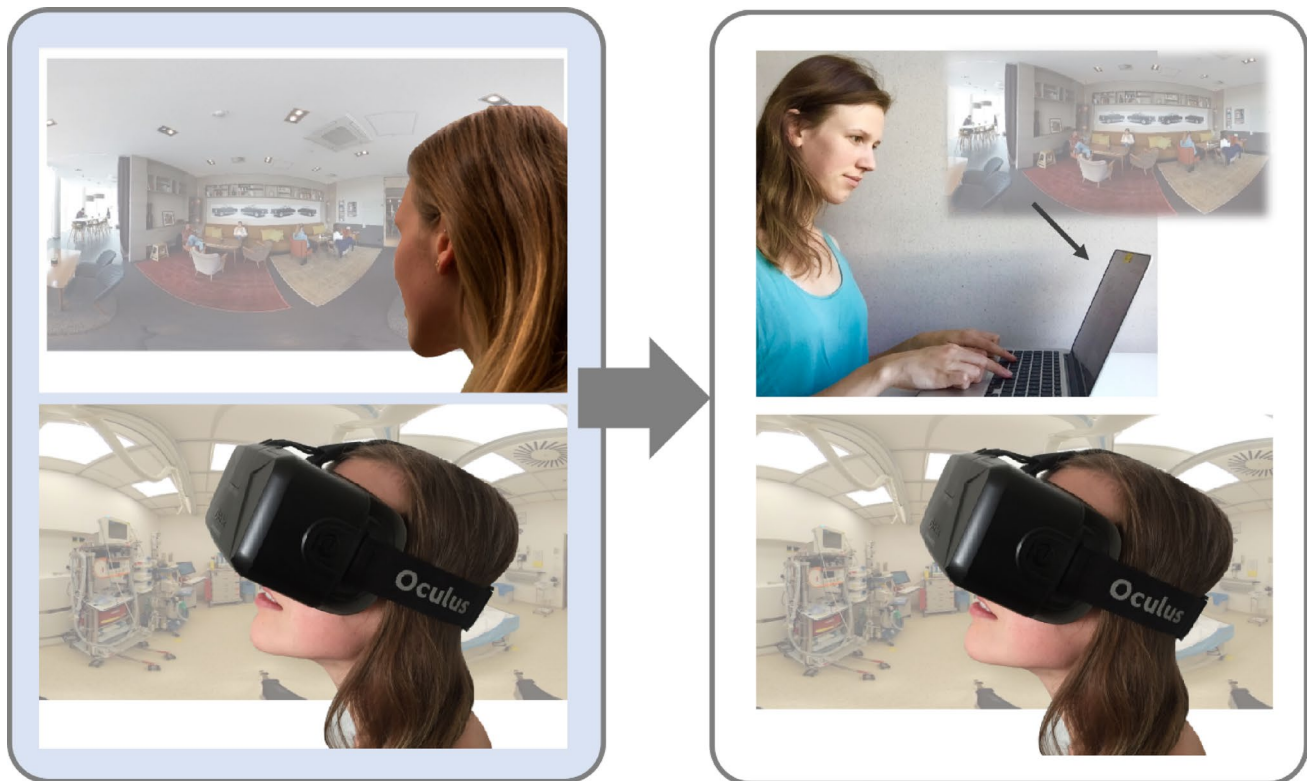


Fig. 1 Schematic illustration of the experimental setup of the encoding (left) and the subsequent retrieval session (right). In the encoding session, all participants rated emotional VR and 2D scenes in terms of valence and arousal (stimuli derived from the luVRe database). The VR scenes were presented via a head-mounted display. The 2D scenes

were large monoscopic video projections. In the subsequent retrieval session, all participants underwent an unannounced recognition task. Half of the participants were presented with the stimuli using a laptop screen, while the other half completed the recognition task in VR

mode conditions. More precisely, we distributed the 16 negative, 16 neutral and 16 positive scenes (as determined in the pilot study, https://osf.io/827jt/?view_only=cc2511cc36ff468fb2c9d8ec74248643) equally between the two encoding presentation modes. However, we would like to put emphasis on the relevance of the individual emotional experience when investigating emotion-related memory processes (Kensinger 2009) and thus consider the valence of the stimuli subjectively perceived by each participant as more important in the main experiment.

2.4 Procedure and task

The encoding session and the retrieval session were conducted on two consecutive days with an approximate time interval of 24 h in between. While prior studies, such as Schöne et al. (2019b), used a 48 h delay, our study differed methodologically; instead of a single immersive event (a VR motorcycle ride), our participants encoded 48 distinct scenes of varying emotional valences. Given the increased cognitive demand of recalling multiple unrelated scenes, we opted for a shorter retention interval (24 h) to facilitate recognition. This adjustment aimed to optimize conditions to

detect a potential VR superiority effect. The participants and the experimenter in the encoding session were blind to the hypotheses, as neither the encoding session nor the retrieval session was announced as such. Two different experimenters conducted the recognition task of the retrieval session, who were not blinded to the hypotheses to be able to fully debrief the participants after the study. Prior to the start of both sessions, all participants gave written informed consent.

Before testing, participants were informed that the study was divided into two sessions on two consecutive days. At the beginning of the encoding session, participants completed demographic questions. Afterwards, they were seated on a swivel chair and watched 48 scenes (24 VR and 24 2D scenes). Participants were tasked to explore the scenes and evaluate them regarding their emotional perception to ensure incidental learning. The experimenter explained that they would watch some scenes in VR and some via a large video projection. Participants rated each scene in terms of *Subjective Valence* and *Subjective Arousal*, using the digital 5-point SAM scales. This approach allows us to disentangle how each participant perceived each stimulus (Kensinger 2009; Zaki and Ochsner 2009). At the end of the encoding session, participants were reminded about the upcoming

second session on the following day and were asked not to consume alcohol or other drugs until then. The experimenter explained that the participants would evaluate additional scenes in the second session.

In the retrieval session on the following day, participants were first informed about the upcoming recognition task. They were instructed to categorize 96 presented static images as “old” or “new” as precisely and as fast as possible by button press in a forced choice old-new judgment task. While an old response indicated that the static image was recognized from a scene presented in the encoding session, a new response indicated that the static image was not recognized from a scene presented in the encoding session on the previous day. In case they were unsure of the answer, they were asked to respond intuitively. The experimenter specifically asked the participants whether they had expected a memory task, which none of them did. Prior to the start of the recognition task, participants completed four training trials with non-experimental material.

Each trial started with a screen reminding the participants to place their fingers on the marked keys. After a blank screen and a 1000 ms fixation cross, a static image was depicted for a maximum duration of 3000 ms. In case participants did not respond within the stimulus presentation interval, a blank screen was shown after the static image presentation. The blank screen was introduced to standardize stimulus viewing times. If a participant did not respond within the fixed presentation window, the blank screen allowed them to provide their response without additional stimulus viewing, which ensured uniform exposure duration across trials. After participants categorized a static image as old or new, they also rated how confident they were in the respective old-new judgment. The 4-point recognition confidence scale ranged from “very unconfident” (1, German: “sehr unsicher”) to “very confident” (4, German: “sehr sicher”). At the end of the experiment, the experimenter thanked the participants, issued the monetary reward and debriefed them about the goal of the experiment. Due to a programming error in the VR recognition task, the presentation of the last static image was missing for 21 participants in the VR retrieval group. At the end of the retrieval session, all participants received a full debriefing about the study’s objectives and hypotheses.

2.5 Data analyses

In the following two sections, we present the analyses used to evaluate our hypotheses, each with respect to recognition performance and recognition confidence. All statistical analyses were performed using R Studio (version: 1.3.959, R version: 3.6.3) (RStudio Team 2020). Standard p -values

of 0.05 were used as the cut-off for statistical testing. Note that we did not exclude any data points as outliers.

2.6 Analyzing presentation mode congruency effects

To evaluate the recognition performance as a function of *Encoding Presentation Mode* and *Retrieval Context* (VR, 2D), we used the sensitivity measure (d'), which accounts for the participant’s response bias driven by non-experimentally controlled influences (Nevin 1969). For each participant, each trial of the recognition task was scored as a hit, correct rejection, miss, or false alarm. We calculated sums of hits, correct rejections, misses, and false alarms per combination of *Participant*, *Encoding Presentation Mode* (VR, 2D) and *Retrieval Context* (congruent, incongruent). The hit rate was the number of correct old-judgments divided by the number of target stimuli in the respective condition. The false alarm rate was calculated as the number of incorrect old-judgments divided by the number of distractor stimuli in the respective condition. The sensitivity (d') was calculated according to the signal detection theory (Green & Swets 1966) with $d' = z(\text{hit rate}) - z(\text{false alarm rate})$. A loglinear correction was applied to hit and false alarm rates in order to also calculate the sensitivity from 100% hit or false alarm rates (Hautus 1995). The effects of *Encoding Presentation Mode* (VR, 2D) and *Retrieval Context* (congruent, incongruent) on the sensitivity were modeled using the `lme()` function from the `nlme` R package (Pinheiro et al. 2019). The linear mixed-effects model included the fixed effects *Encoding Presentation Mode* and *Retrieval Context* as main and interaction effects. The predictor *Participant* was added as a random intercept to account for interindividual differences.

We generated a second linear mixed-effects model to analyze the recognition confidence, using the same two predictors and the same random effect as in the sensitivity model. We aggregated confidence ratings per participant and condition, and used mean confidence as continuous outcome measure in the model. The analysis of the recognition confidence was originally planned for all target trials but was slightly adapted and restricted to hits to focus on the confidence for successful memory recall. Based on the parameters of the sensitivity and the recognition confidence model, we computed two analyses of variance (Type III), using the `anova()` function from the `stats` R package (R Core Team 2020). We followed up significant effects of both models with two-tailed contrasts with model-based standard errors, using the `glht()` function with Bonferroni correction from the `multcomp` R package (Hothorn et al. 2008).

2.7 Analyzing effects involving the individually rated emotional content

To evaluate whether the encoding presentation mode and the emotional content interact with respect to memory performance, we considered the *Subjective Valence* and *Subjective Arousal* in the second part of the analyses. Note that we have planned further analyses (e.g., to investigate the influence of the pre-rated emotional valence of the scenes, as determined in a pilot study based on ratings of other participants). Nonetheless, we refrain from testing them, because the valence ratings were largely inconsistent between this and the pilot study, as revealed by a Cohen's $\kappa=0.05$ between *Predetermined* and *Subjective Valence*. Instead, we focused on the effects of *Subjective Valence* and *Arousal*, as indicated by the main study sample, which was assumed to be relevant for the individual memory processes.

Since the participants only indicated their emotional ratings for the presented scenes of the encoding phase, but not for the distractor stimuli, it was not possible to determine the false alarm rates for the different emotional categories. Therefore, we could not apply signal detection theory and utilize the sensitivity measure d' to quantify memory performance in this part of the analyses. Instead, we used the accuracy defined as a binary outcome measure with correct (hits=1) and incorrect (misses=0) responses in single target trials only. Based on the individual valence and arousal ratings of both retrieval groups, we created a generalized linear mixed-effects model with the fixed effects *Encoding Presentation Mode* (2D, VR), *Subjective Valence* (negative, neutral, positive) and *Subjective Arousal* (continuous) to evaluate the interaction between encoding presentation mode and emotional valence. The model was implemented using the `glmer()` function from the `lme4` R package (Bates et al. 2015). The model also included *Participant* and *Stimulus* as random intercepts. *Participant* was included as a random intercept in the analyses because the inter-individual variability was expected to influence the results. We added *Stimulus* as an additional random effect in the models, for which variability between scenes was also expected (e.g., for different emotional valences). That is, we included *Stimulus* as an additional random effect when valence served as predictor.

We assigned the values 1 & 2, value 3 and values 4 & 5 of the 5-point SAM valence scale to the levels negative, neutral, and positive of the factor *Subjective Valence*, respectively. The 5-point Likert scale was recoded into a 3-point scale to ensure a sufficient number of trials per condition and enhance statistical power. This approach preserved distinctions between negative, neutral, and positive valence categories. In addition, we applied a polynomial contrast for *Subjective Valence*, such that the model also included

a quadratic term for the comparison of the two emotional levels (negative and positive) with the neutral level in line with the emotional enhancement effect.

The covariate *Subjective Arousal* was group-mean centered to account for differences in arousal ratings between the three valence conditions. This approach ensured that the effect of valence on recognition performance and confidence was not confounded by variations in arousal. While it was likely that arousal ratings differed between negative, neutral, and positive stimuli, our hypotheses focused specifically on effects of valence (and of presentation mode/retrieval context) but not of arousal. As a result, we did not assess arousal as an independent predictor.

To account for the unbalanced data in the *Subjective Valence* and *Subjective Arousal* ratings, we conducted Type II Wald chi-square tests on the parameters of the accuracy model including these two predictors. This was done using the `Anova()` function from the `car` R package (Fox and Weisberg 2019). We followed up significant effects with two-tailed contrasts with model-based standard errors, using the `glht()` function with Bonferroni correction from the `multcomp` R package (Hothorn et al. 2008).

Using the same three predictors and random effects as in the accuracy model, we computed a cumulative link mixed-effects model to study effects on the recognition confidence for trials scored as hits. As this model was based on single trials, recognition confidence was included as an ordinal outcome variable (levels=1, 2, 3, 4) in the model. It was implemented using the `clmm()` function from the `ordinal` R package. Follow-up comparisons were conducted using reference coding, with the neutral valence condition specified as the reference level.

3 Results

We hypothesized that scenes encoded in one presentation mode would generally be recognized better and with higher confidence when presented in the congruent presentation mode during the retrieval phase (context congruency effect, Hypothesis 1). In addition, we hypothesized that this context congruency effect would be lower for the VR- than for the 2D-encoded scenes because transfer costs could be lower for scenes incidentally learned in VR (Hypothesis 2). Moreover, we expected an emotional enhancement effect and therefore hypothesized that scenes subjectively perceived as emotional (negative, positive) would be recognized better and with higher confidence, particularly for scenes viewed in VR (Hypothesis 3). These hypotheses are evaluated in the following sections.

3.1 Effects of presentation mode on recognition performance and confidence

Figure 2 shows the sensitivity (d') as a function of encoding presentation mode (VR, 2D) and retrieval context (congruent, incongruent to encoding). For 2D-encoded scenes, recognition sensitivity was descriptively higher in the congruent context ($M=2.06$, $SD=0.49$) than in the incongruent retrieval context ($M=0.64$, $SD=0.82$). For VR-encoded scenes, this pattern was reversed: sensitivity was higher in the incongruent ($M=1.67$, $SD=0.43$) than in the congruent retrieval context ($M=0.97$, $SD=1.87$). A linear mixed-effects model on the mean sensitivity revealed a significant main effect of retrieval context, $F(1, 58)=6.97$, $p=0.011$, $\eta_p^2 = 0.11$, as well as a significant interaction with encoding presentation mode, $F(1, 58)=19.14$, $p<0.001$, $\eta_p^2 = 0.25$ (see also Appendix Table A2). Follow-up contrasts indicated that the congruency effect (i.e., better recognition for congruent vs. incongruent retrieval contexts) was only present for 2D-encoded scenes, $z=5.11$, $p<0.001$, $d=0.95$. In contrast, for VR-encoded scenes, recognition performance was higher in the incongruent condition, $z=2.50$, $p=0.025$, $d=1.01$. No significant main effect of encoding presentation mode emerged, $F(1, 58)=0.06$, $p=0.812$.

A similar pattern was observed for the mean recognition confidence of hit trials. A linear mixed-effects model revealed a significant main effect of retrieval context, $F(1, 58)=14.70$, $p<0.001$, $\eta_p^2 = 0.20$, as well as a significant interaction with encoding presentation mode, $F(1,$

$58)=8.13$, $p=0.006$, $\eta_p^2 = 0.12$ (see also Appendix Table A3). For 2D-encoded scenes, confidence was higher in the congruent than in the incongruent condition, $z=4.60$, $p<0.001$, $d=0.51$, while no such difference was found to be significant for VR-encoded scenes, $z=0.08$, $p>0.999$. No significant main effect of encoding presentation mode emerged, $F(1, 58)=0.22$, $p=0.638$.

Taken together, scenes encoded in a 2D presentation mode were recognized both more accurately and with higher confidence when the retrieval context was congruent compared to incongruent. In contrast, this congruency benefit was not observed for scenes encoded in VR (partial support of Hypothesis 1). Notably, recognition performance and confidence for VR-encoded scenes did not significantly decrease in the incongruent condition, suggesting a reduced sensitivity to contextual congruence (support of Hypothesis 2). Overall, recognition performance and confidence were similar across presentation modes (VR sensitivity: $M=1.32$, $SD=1.39$, confidence: $M=3.07$, $SD=0.51$; 2D sensitivity: $M=1.35$, $SD=0.98$, confidence: $M=3.10$, $SD=0.45$).

3.2 Effects of subjectively perceived emotion

Figure 3 depicts the mean accuracy in target trials and the recognition confidence in hit trials as a function of encoding presentation mode and subjective valence. With respect to accuracy, the emotional enhancement effect was descriptively evident for 2D-encoded scenes (negative: $M=0.67$, $SD=0.25$; neutral: $M=0.63$, $SD=0.24$; positive: $M=0.67$,

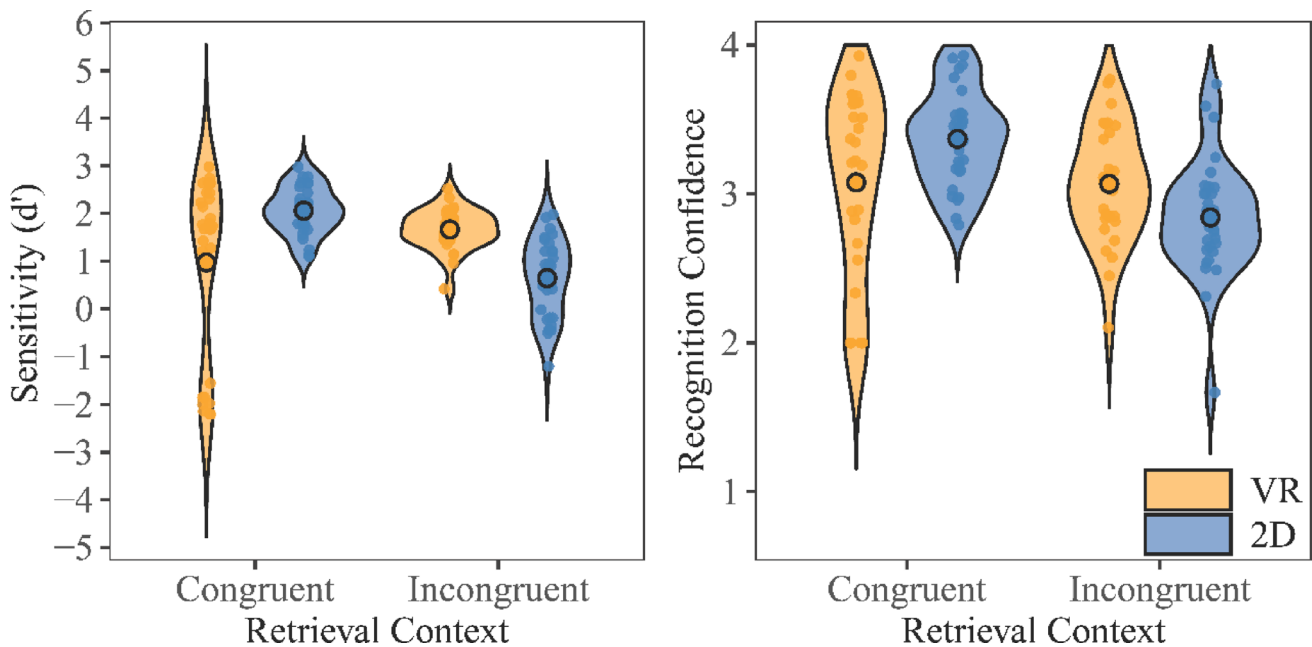


Fig. 2 Sensitivity as measured by d' (left) and confidence (right) in the recognition task, both as a function of the encoding presentation mode (color) and the retrieval context (x-axis). Orange: VR-encoding. Blue:

2D-encoding. Violin plots represent the data distribution; larger black outlined dots show condition means; individual participant means are shown as smaller filled dots

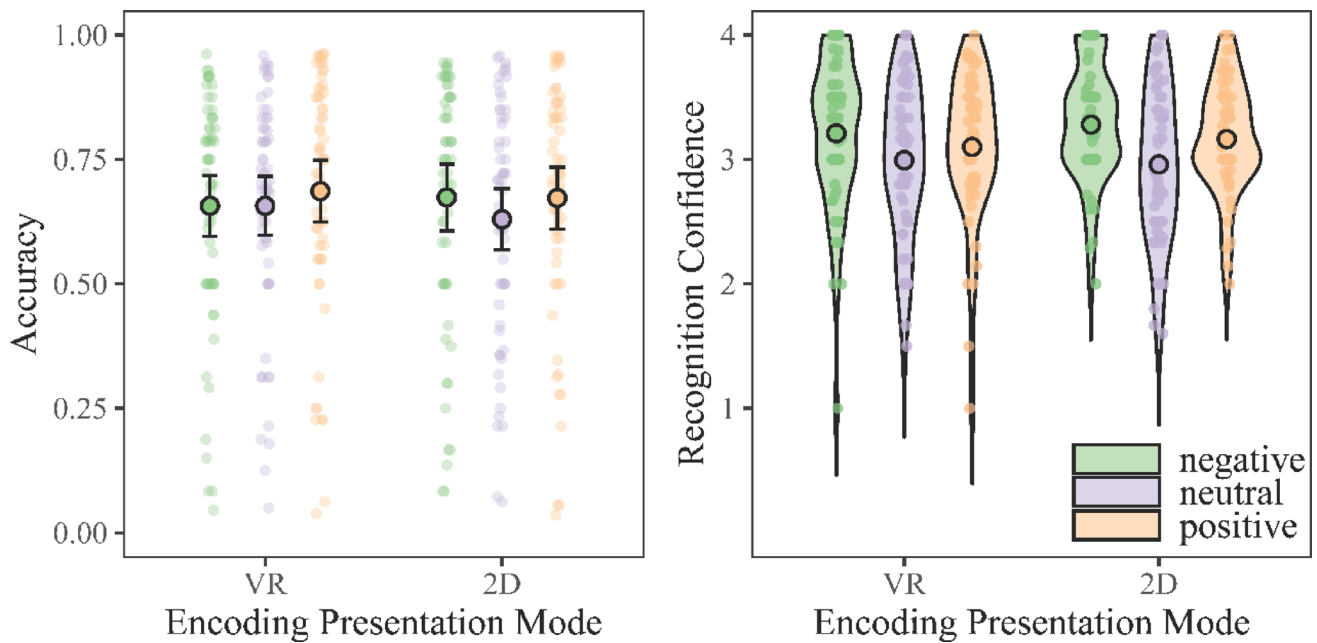


Fig. 3 Accuracy (left) in target trials and confidence in hit trials (right) of the recognition task, both as a function of the individually rated emotional valence (color) and the encoding presentation mode (x-axis). Green: negative valence. Purple: neutral valence. Yellow:

positive valence. Error bars in left panel show 95% confidence intervals of the individual means. Black outlined dots show condition means in right panel. In both panels, individual participant means are shown as smaller filled dots

$SD=0.24$) and partially for VR-encoded scenes (negative: $M=0.66$, $SD=0.24$; neutral: $M=0.66$, $SD=0.23$; positive: $M=0.69$, $SD=0.24$). A generalized linear mixed-effects model on the accuracy in target trials revealed a significant main effect of the subjective valence, $\chi^2(2)=6.10$, $p=0.047$, $OR_{lin} = 1.21$, $OR_{quad} = 1.04$ (see also Appendix Table A4). Follow-up comparisons showed that participants identified targets significantly more accurately when the stimuli were rated as emotionally positive compared to neutral, $z=2.42$, $p=0.047$, $OR=1.26$, compatible with the emotional enhancement effect. No other pairwise comparisons reached significance, all $p>0.090$. Contrary to our hypothesis, neither the main effect of the encoding presentation mode, $\chi^2(1)=0.43$, $p=0.431$, nor the interaction with subjective valence were significant, $\chi^2(2)=3.40$, $p=0.183$.

Regarding recognition confidence, participants reported higher confidence for emotional than for neutral scenes – both for VR-encoded (negative: $M=3.02$, $SD=0.53$; neutral: $M=2.89$, $SD=0.55$; positive: $M=2.92$, $SD=0.56$) and 2D-encoded scenes (negative: $M=3.12$, $SD=0.52$; neutral: $M=2.87$, $SD=0.55$; positive: $M=2.93$, $SD=0.45$). The cumulative link mixed-effects model revealed a significant main effect of subjective valence (quadratic term), $\beta=0.27$, $SE=0.11$, $z=2.37$, $p=0.018$, $OR=1.31$ (see also Appendix Table A5). Follow-up contrasts confirmed significantly higher confidence ratings for positive than for neutral, $z=3.62$, $p<0.001$, $OR=1.51$, and for negative valence, $z=2.22$, $p=0.027$, $OR=1.37$.

In sum, subjective emotional valence significantly influenced both recognition accuracy and confidence. Scenes perceived as emotionally positive were better remembered and recognized with greater confidence than neutral scenes, consistent with the emotional enhancement effect. This pattern did not significantly differ between encoding presentation modes, contrary to our expectation (partial support of Hypothesis 3).

3.3 Exploratory analyses

Additionally, we explored whether scenes viewed in VR were rated as more emotional (negative, positive) compared to the same scenes viewed in 2D during the encoding phase. This could potentially explain why the emotional enhancement effect was not significantly different between the two encoding presentation modes. For instance, if participants tended to perceive VR scenes as more emotional (negative, positive) than the same scenes in 2D, the emotional enhancement effect might overshadow the effect of the encoding presentation mode on memory. To examine this exploratory hypothesis statistically, we fitted a generalized linear mixed-effects model, which included the binary outcome emotional rating (1=emotional, 0=neutral), the fixed effect *Encoding Presentation Mode* (2D, VR) as well as *Participant* and *Stimulus* as random intercepts to account for differences between participants and stimuli. Type II Wald chi-square tests on the parameters of

the model confirmed that VR scenes were more likely to be rated as emotional (negative, positive) than when the same scenes were viewed as 2D video during the encoding phase, $\chi^2(1)=9.13$, $p=0.003$, $OR=1.24$. Consequently, the emotional enhancement effect could have mediated the effect of the encoding presentation mode on recognition performance and confidence. For the sake of completeness, we report the mean arousal ratings during VR- ($M=2.23$, $SD=0.46$) and 2D-encoding ($M=1.98$, $SD=0.44$). However, we refrain from statistical testing because the experiment did not aim to evaluate the effect of arousal on memory processes and was therefore not designed as such (i.e., we did not aim to counterbalance different levels of arousal across the encoding contexts).

4 Discussion

In the current study, all participants first underwent an encoding phase, in which they viewed emotional (positive, negative) and neutral scenes in both VR and 2D and rated each scene with respect to emotional valence and arousal. They completed an unannounced recognition task one day later. Half of the participants completed the recognition task in VR, while the other half completed it on a 2D screen. We measured the recognition performance and recognition confidence for each emotional category, encoding context (VR, 2D), and retrieval context (congruent, incongruent with encoding). Our results showed that a context congruency effect occurred for scenes that participants encoded in 2D. We observed that for scenes encoded in VR, the congruency with the encoding presentation mode played a minor role, as participants managed to recognize the scenes in an incongruent retrieval context. For correctly recognized scenes (hits), participants were significantly more confident in the recognition of 2D-encoded scenes in the congruent than in the incongruent retrieval context. For VR-encoded scenes, the recognition confidence in hit trials did not vary significantly between the retrieval contexts. Furthermore, we found that participants recognized VR- and 2D-encoded scenes, which they had previously rated as emotional, better and more confidently than neutral scenes, compatible with the emotional enhancement effect. In summary, our data support the idea that emotional content enhances recognition processes, irrespective of the presentation mode. However, it is crucial to note that the encoding of scenes in 2D enhanced memory processes on one hand but introduced a stronger context congruency effect on the other hand, while the encoding of scenes in VR depended less on context congruency.

4.1 Context congruency

Numerous studies have reported context congruency effects on both memory recall and recognition (Godden and Baddeley 1975, 1980; Murnane et al. 1999; Murnane and Phelps 1993; Smith and Manzano 2010; Smith and Vela 2001). However, the current study showed that congruency in a (2D) presentation mode can also be effective as a contextual aid for incidental scene recognition. Despite employing a large projection for the implementation of 2D-encoding, the presentation mode featured a distinct spatial scene layout. Specifically, a salient element within the 2D scene might have been displayed at a particular spatial position, for instance, in the bottom left corner. This spatial arrangement could thus have allowed participants to utilize the scene layout as a source of information during the 2D recognition task, potentially enhancing memory retrieval in congruent 2D conditions. In contrast, the scene layout of a VR scene was much more complex because the content was distributed across three dimensions and a complete 360° spectrum. Participants who encoded a 2D scene might have had difficulty recognizing it in VR due to the challenge of mapping the learned 2D scene layout to the “new” three-dimensional VR scene, which could potentially explain the substantial transfer costs for scenes encoded in 2D but retrieved in VR. Even if they correctly recognized a 2D-encoded scene in the VR recognition task, participants indicated significantly lower recognition confidence. This highlights that the transfer costs for content incidentally learned in 2D expanded to the subjective experience during retrieval.

When participants encoded a VR scene, they had to engage a more active exploration behavior to decipher the scene than was the case for the 2D presentation mode. In fact, the experimenters instructed the participants explicitly to turn their heads to explore the full scene. The rather active and self-referential exploration of VR scenes (Gupta et al. 2025; S. A. Smith 2019) could have promoted the consolidation of the spatial scene content, resulting in lower dependency on context congruency when retrieving scenes encoded in VR. In line with this interpretation, recent research (Johnsdorf et al. 2025) suggests that the encoding demands imposed by three-dimensional content, particularly the need for active, self-directed exploration, are crucial for enhanced memory performance, compared to two-dimensional content – presumably independent of whether it is presented within a VR environment or on a conventional screen as in the present study. At the same time, free exploration in three-dimensional VR environments may introduce additional inter-individual variability, as participants differ in their exploration strategies, which

could explain the substantial variability observed particularly in the congruent VR encoding and retrieval condition in this study.

The hypothesis that active scene exploration is crucial for a VR-encoding advantage in memory to occur would be compatible with the study by Johnsdorf et al. (2023b) and Kulke and Pasqualetto (2024), in which no significant superiority effect emerged for a passive VR-based learning task of a spatial arrangement of buildings and an electroencephalography training, respectively. Notably, our data also showed a higher recognition performance for VR-encoded scenes in the incongruent 2D recognition task compared to the congruent one. This difference might be attributed to a generally lower difficulty in evaluating stimuli in the 2D recognition task. However, even among correctly recognized VR-encoded scenes, participants reported similar levels of recognition confidence for both retrieval contexts. Consequently, the subjective experience during the recognition of VR-encoded scenes remained largely unaffected by the retrieval context, indicating that the VR presentation mode serves less as a contextual aid for incidental scene recognition when compared to the 2D presentation mode. Instead, our findings suggest that the recognition of learning content encoded in VR exhibited remarkable resilience against changes in the retrieval context.

While recognition confidence for VR-encoded scenes appeared stable across contexts, our findings still cannot distinguish whether this reflects genuinely a more vivid memory representation or results from potentially mediating factors such as a greater sense of presence. In line with previous studies (Johnsdorf et al. 2023a; Kisker et al. 2019b), it is possible that VR-encoding directly enhanced vivid and recollective processing. At the same time, higher immersion such as in VR exhibits a stronger feeling of presence than lower immersion (Diemer et al. 2015; Kuhne et al. 2023), which may indirectly facilitate mnemonic processes (Makowski et al. 2017). However, findings by Cadet and Chainay (2020) suggest that while immersion in VR increases presence and interacts with emotional experience, it does not directly enhance memory performance. Instead, they found that the emotional enhancement of memory emerged primarily in immersive VR conditions, implying that emotion – rather than presence or immersion alone – might play a more critical role in modulating memory within VR. Future research should explicitly investigate these possibilities, for instance, by varying different levels of immersion within VR and disentangling the effects of immersion, emotion, and presence on memory outcomes.

The implications of the study highlight the potential challenges associated with transferring content learned in 2D to diverse (three-dimensional) contexts. In educational and therapeutic settings, learning often occurs in settings that

differ from the situations in which the acquired knowledge or skills are later applied. While education broadly targets understanding, reasoning, and domain-specific knowledge, and therapy aims at symptom reduction, functional improvement, and emotion regulation, successful generalization across contexts partially relies on the retrieval of previously acquired experiences (Kredlow et al. 2018; Lane et al. 2015; Sarfan et al. 2023). In this regard, susceptibility to transfer costs may pose limitations for learning content that is bound to its original presentation context. Therapeutic interventions, for example, often aim to support the application of coping strategies or emotional insights beyond the treatment setting, which may depend in part on the robustness of memory representations (Harvey et al. 2016; Lane et al. 2015). Our study suggests that VR settings could be more suitable in such cases due to higher emotional engagement and a lower mnemonic dependency on context congruency. This is particularly relevant in the context of VR exposure therapy, a widely used intervention for treating anxiety and post-traumatic stress disorders, where patients are gradually exposed to anxiety-provoking stimuli in immersive VR environments (Carl et al. 2019; Diemer et al. 2014; Gonçalves et al. 2012; Kothgassner et al. 2019; Schröder et al. 2023; Shahid et al. 2024; Wong et al. 2023; Wu et al. 2021). The generalization of coping strategies from the virtual to the real world, which is a core aim of VR-based exposure therapy, will benefit from less context-independent memory traces. Also, VR-based emotion regulation training, which places participants in emotionally evocative but controlled environments to practice techniques such as reappraisal or emotional distancing, have shown to be effective (Montana et al. 2020; Roa and Rodríguez 2024; Yu et al. 2023). Thus, successful application of VR therapy could be partially related to robust memory formation and may therefore foster the transfer to the real world (Bonnail et al. 2024; Kisker et al. 2020; Martarelli et al. 2023). In addition, our study also shed light on the subjective confidence in memory retrieval. In the context of psychological disorders, successful memory retrieval accompanied by a robust level of confidence could align well with effective therapeutic outcomes. In contrast, the specificity of incidentally learned content to the 2D presentation mode context, such as recognizing scenes from a movie seen in a cinema, may be better suited for retrieval in a similar 2D presentation mode, such as when watching at home on a TV screen. While the latter case may have limited relevance in many learning contexts, the overall message is clear: designing learning settings that maximize the probability of successful transfer to new situations requires careful consideration of the presentation mode. VR environments may offer advantages in creating robust memories of dynamic emotional scenes applicable

across diverse (real-world) scenarios, making them a valuable tool in educational and therapeutic contexts.

4.2 Emotional enhancement effect

The subjective emotional ratings of the scenes were significantly related to the participants' performance and confidence in recognizing them the following day, consistent with the emotional enhancement effect (Adelman and Estes 2013; Dolcos and Cabeza 2002; Harris and Pashler 2005; Kensinger 2009). Specifically, participants recognized scenes that they had previously evaluated with a positive valence with greater accuracy and confidence than scenes they had assessed neutrally. Negative scenes were not recognized significantly better than neutral scenes in our study. Hence, the emotional enhancement effect with respect to the (objective) performance measure was only observed for the positive emotional valence. This finding is in line with a number of studies, such as the study by Erk et al. (2003), who found positive words to be better memorized than negative or neutral words, and the study by Chainay et al. (2012), who showed that incidentally learned positive pictures could be better recognized than negative or neutral stimuli. Nonetheless, for correctly recognized scenes, participants felt substantially more confident about their recognition judgment for both positive and negative scenes than for neutral scenes. Thus, the emotional enhancement effect occurred for both emotional valences on the subjective measure. In sum, this shows that the subjectively perceived emotional content is an important predictor for successful memory retrieval.

Contrary to our expectations, the emotional enhancement effect did not significantly vary between the two encoding presentation modes. However, participants rated VR scenes significantly more often as emotional (negative, positive valence) than 2D videos. This suggests that evoking an emotional response may be more likely in VR than in a 2D format, potentially due to higher immersion in VR (Kuhne et al. 2023; Pavic et al. 2023). Increased immersion may intensify emotional responses by creating a greater sense of presence. This idea aligns with findings from Cadet and Chainay (2020), who demonstrated that immersion through the use of a head-mounted display was associated with increased emotional responses and a greater sense of presence. Interestingly, they found no direct effect of immersion on memory performance, while the emotional enhancement effect on memory only emerged in the VR but not in the computer screen condition. It should be mentioned that immersion as a possible influence on the emotional enhancement effect may apply to realistic scenes but potentially not to all stimuli presented in VR. Linguistic stimuli, for example, tend to be more stable in their emotional

evaluation (Rocabado and Duñabeitia, 2024). Nevertheless, our data do not provide strong evidence that VR directly enhances encoding depth beyond its effects on emotional perception. Thus, further investigation is needed to determine whether VR enhanced memory encoding directly, or if it primarily enhanced emotional perception (Gupta et al. 2025; Riva et al. 2007), which in turn improved memory retrieval. Similarly, the richer sensory input and the need for active exploration in VR (van Helvoort et al. 2020) could have imposed a higher cognitive load (i.e., the amount of mental effort required to process information) compared to the more passive experience of 2D, which could have enhanced emotional engagement at the same time. Finally, habituation (i.e., becoming accustomed to stimuli) could have been relevant in this regard. More precisely, participants unfamiliar with VR may have experienced novelty effects with elevated emotional responses specific to this presentation mode.

Future studies should aim to disentangle whether the emotional enhancement effect observed in our study is directly driven by a deeper encoding or primarily intensified emotional perception at the time of encoding. For instance, by controlling perceived emotional intensity across presentation modes and by including neurophysiological measures (e.g., EEG or skin conductance; Gupta et al. 2022) to assess encoding-related physiological activity beyond behavioral outcomes.

From a practical point of view, it would be beneficial to design learning content in such a way that it elicits a (positive) emotional response, which may be easier in VR, to facilitate successful memory retrieval in the future. Note, however, that in our study we used each participant's subjective emotional ratings for the scenes as a predictor, which varied between the participants and deviated from the ratings obtained from another sample in our pilot study. This demonstrates the need to investigate how (positive) emotions can be reliably elicited for a large population or for specific subsamples.

4.3 Limitations

Although the present study provides valuable insights, it is not without limitations. Firstly, the recognition assessment was limited to a 24 h interval, restricting the generalizability of our results to relatively short retention periods. Future investigations should explore how these findings extend to longer retention intervals (e.g., multiple days; Schöne et al. 2019b), as such durations are often more pertinent in learning contexts. Additionally, our study employed only two presentation modes, namely 2D and VR, with speculative implications for real-world transfer. As some studies suggest similar emotional and mnemonic processes in VR and

real-world conditions (Kisker et al. 2019a; Schöne et al. 2023), but others report differences at least for intentionally learned content (Monaro et al. 2024), future research should test real-world conditions to determine whether incidental emotional memories generated in VR can be easily transferred to real-world scenarios.

Another constraint of the study is the lack of control over how participants visually explored the scenes. In the VR-encoding context, participants freely explored the scenes, such that their perception of the scene depended on their head movements. These movements were likely to differ between encoding and retrieval phases, even within the same participant. This presumably introduced variability in how the visual information was processed across sessions. In the 2D conditions, we attempted to introduce a change in context by presenting the encoding scenes on a large video projection and the retrieval scenes on a standard computer screen. However, the visual layout remained fixed, unlike the self-directed exploration in VR. These differences in scene interaction and visual dynamics may have influenced memory performance and should be carefully controlled or systematically manipulated in future studies.

Further constraints of our study include that we did not use eye-tracking technology or other physiological measures. As a result, we lack information regarding the specific areas or items within the scenes that were encoded. This uncertainty limits the generalizability of our findings to diverse learning contents, such as item recognition. In addition, the use of physiological measures allowing a more objective assessment of emotional engagement (e.g., skin conductance and pupil dilation) could have ensured the reliability of the subjectively perceived emotion and should therefore be added in future studies. Also, we did not assess the feeling of presence or VR experiences in this study, which limits our ability to determine whether increased presence or a novelty effect contributed to participants rating VR scenes as more emotional. Moreover, the factor *Retrieval Context* was a between-subjects variable in our study, which might have reduced the test power for the main effect of the factor as well as for the interaction between the encoding presentation mode and the retrieval context. Finally, although we accounted for individual differences in emotional responses to the scenes by using participants' subjective valence ratings as predictor in our statistical models to derive general conclusions, it remains possible that VR amplifies emotional responses more strongly in some individuals than in others (Wang et al. 2022) and may therefore translate into interindividual differences in memory processes. In this regard, we focused on the dimension of valence and group-mean centered subjective arousal ratings per valence category to isolate the effect of valence, as it was crucial to evaluate the established hypotheses. However, we did

not further explore interindividual variability in emotional responses to VR versus 2D, as this would require a larger sample to yield reliable conclusions. Future studies should explore how interindividual differences in emotional states and traits may interact with memory processes in VR.

4.4 Conclusion

In summary, this study showed that a context congruency effect occurred for scenes learned in 2D, while content learned in VR could be retrieved reliably and confidently in an incongruent retrieval context. Furthermore, the study revealed the important effects of (positive) emotional content on learning success. The findings have practical implications: learning success can be improved by considering context effects and by utilizing (positive) emotional content, the latter possibly being easier in VR.

Appendix

See Tables 1, 2, 3, 4 and 5

Appendix

Table 1 Predetermined Valence category based on pilot study (https://osf.io/827jt/?view_only=cc2511ec36ff468fb2c9d8ec74248643) for all encoding scenes derived from the LuVRe database. A Valence value of 1, 2 and 3 represents a negative, neutral and positive Valence category, respectively

Video	Predetermined Valence
x_1002	3
x_1008	1
x_1011	1
x_1106	1
x_1201	2
x_1307	2
x_1502	1
x_1609	3
x_1613	3
x_1802	1
x_1908	2
x_2003	1
x_0202	2
x_2104	2
x_2105	2
x_2202	2
x_2303	1
x_2306	1
x_2310_2	1
x_2502	1
x_2603	3

Table 1 Predetermined Valence category based on pilot study (https://osf.io/827jt/?view_only=cc2511ec36ff468fb2c9d8ec74248643) for all encoding scenes derived from the LuVRe database. A Valence value of 1, 2 and 3 represents a negative, neutral and positive Valence category, respectively

Video	Predetermined Valence
x_2701	1
x_2704	1
x_2803	3
x_2903	3
x_0309	2
x_3101	3
x_0311	2
x_3301	3
x_3310	2
x_3602	3
x_3611	3
x_3801	3
x_3901	2
x_0401	1
x_4209	3
x_4215	2
x_4219	3
x_4222	3
x_4226	2
x_4303	1
x_4402	2
x_4502	2
x_4602	2
x_4702	3
x_4703	1
x_0702	3
x_0802	1

Table 2 Parameter estimates of the linear mixed-effects models predicting sensitivity (d') as a function of encoding presentation mode and retrieval context

Predictor	beta	SE	df	t	p
Encoding PM	-0.016	0.07	58	0.24	0.812
Retrieval C	0.180	0.07	58	2.64	0.011
Encoding PM × Retrieval C	-0.527	0.12	58	4.37	<0.001

Table 3 Parameter estimates of the linear mixed-effects models predicting mean recognition confidence as a function of encoding presentation mode and retrieval context

Predictor	beta	SE	df	t	p
Encoding PM	-0.016	0.04	58	0.47	0.812
Retrieval C	0.135	0.04	58	3.83	<0.001
Encoding PM × Retrieval C	-0.130	0.05	58	2.85	0.006

Table 4 Parameter estimates of the generalized linear mixed-effects model predicting recognition accuracy (correct vs. incorrect) in target trials based on encoding presentation Mode, subjective Valence (linear, quadratic term) and group-mean centered arousal

Predictor	beta	SE	z	p
Encoding PM	-0.033	0.09	0.38	0.705
Valence (lin)	0.190	0.11	1.67	0.096
Valence (quad)	0.039	0.11	0.36	0.717
Arousal	0.026	0.05	0.52	0.602
Encoding PM × Valence (lin)	-0.196	0.16	1.26	0.209
Encoding PM × Valence (quad)	0.224	0.14	1.54	0.123

Table 5 Parameter estimates of the cumulative link mixed-effects model predicting recognition confidence in hit trials based on encoding presentation Mode, subjective Valence (linear, quadratic term) and group-mean centered arousal

Predictor	beta	SE	z	p
Encoding PM	-0.031	0.10	0.33	0.745
Valence (lin)	0.014	0.13	0.10	0.918
Valence (quad)	0.271	0.11	2.37	0.018
Arousal	0.122	0.05	2.21	0.027
Encoding PM × Valence (lin)	0.118	0.16	0.71	0.476
Encoding PM × Valence (quad)	0.045	0.16	0.29	0.774

Acknowledgements We thank the Fachgruppe Psychologie der Georg-August-Universität Goettingen and Leibniz ScienceCampus “Primate Cognition” for financial support and Jana Znaniewitz, Lena Brümmer, Mareike Spengler for their assistance with data collection.

Author contributions M. Wessels, C. Valuch and L. Kulke developed the study concept and contributed to the study design. Testing and data collection were performed by M. Wessels and research assistants, supervised by C. Valuch and L. Kulke. M. Wessels performed the data analysis under the supervision of C. Valuch and L. Kulke. M. Wessels, C. Valuch and L. Kulke interpreted the data. M. Wessels drafted the manuscript, and C. Valuch, L. Kulke and A. Schacht provided critical revisions. All authors approved the final version of the manuscript for submission.

Funding Open Access funding enabled and organized by Projekt DEAL. Fachgruppe Psychologie der Georg-August-Universität Goettingen, Leibniz ScienceCampus “Primate Cognition”

Data availability All data are shared openly alongside the manuscript (Open Science Framework, [https://osf.io/3vweg/?view_only=900666375c6a4c299e06e4ab4066ca3c], Identifier: <https://doi.org/10.17605/OSF.IO/3VWEG>).

Declarations

Conflict of interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate

if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Adelman JS, Estes Z (2013) Emotion and memory: a recognition advantage for positive and negative words independent of arousal. *Cognition* 129(3):530–535. <https://doi.org/10.1016/j.cognition.2013.08.014>
- Allcoat D, von Mühlhagen A (2018) Learning in virtual reality: effects on performance emotion and engagement. *Res Learn Technol*. <https://doi.org/10.25304/rlt.v26.2140>
- Bates D, Mächler M, Bolker B, Walker S (2015) Fitting linear mixed-effects models using lme4. *J Stat Softw* 67(1):1–48. <https://doi.org/10.18637/jss.v067.i01>
- Bonnail, E., Frommel, J., Lecolinet, E., Huron, S., & Gugenheimer, J. (2024). Was it real or virtual? Confirming the occurrence and explaining causes of memory source confusion between reality and virtual reality. *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1–17. <https://doi.org/10.1145/3613904.3641992>
- Bradley MM, Lang PJ (1994) Measuring emotion: the self-assessment manikin and the semantic differential. *J Behav Ther Exp Psychiatry* 25(1):49–59. [https://doi.org/10.1016/0005-7916\(94\)90063-9](https://doi.org/10.1016/0005-7916(94)90063-9)
- Cadet LB, Chainay H (2020) Memory of virtual experiences: role of immersion, emotion and sense of presence. *Int J Hum Comput Stud* 144:102506. <https://doi.org/10.1016/j.ijhcs.2020.102506>
- Cadet LB, Reynaud E, Chainay H (2021) Memory for a virtual reality experience in children and adults according to image quality, emotion, and sense of presence. *Virtual Reality*. <https://doi.org/10.1007/s10055-021-00537-y>
- Carl E, Stein AT, Levihn-Coon A, Pogue JR, Rothbaum B, Emmelkamp P, Asmundson GJG, Carlbring P, Powers MB (2019) Virtual reality exposure therapy for anxiety and related disorders: a meta-analysis of randomized controlled trials. *J Anxiety Disord* 61:27–36. <https://doi.org/10.1016/j.janxdis.2018.08.003>
- Chainay H, Michael GA, Vert-pré M, Landré L, Plasson A (2012) Emotional enhancement of immediate memory: positive pictorial stimuli are better recognized than neutral or negative pictorial stimuli. *Adv Cogn Psychol* 8(3):255–266. <https://doi.org/10.2478/v10053-008-0121-1>
- Chirico A, Gaggioli A (2019) When virtual feels real: comparing emotional responses and presence in virtual and natural environments. *Cyberpsychol Behav Soc Netw* 22(3):220–226. <https://doi.org/10.1089/cyber.2018.0393>
- Choi Y, Kim K, Moon S, Jung G, Cha J-H, Yim H (2025) Context-dependent memory in the real world: the role of frequency and context dwell time. *Front Psychol*. <https://doi.org/10.3389/fpsyg.2024.1489039>
- Clay CJ, Budde JR, Hoang AQ, Gushchyan A (2024) An evaluation of the effectiveness of immersive virtual reality training in non-specialized medical procedures for caregivers and students: a brief literature review. *Front Virtual Real*. <https://doi.org/10.3389/frv.2024.1402093>
- Conway MA (2001) Sensory–perceptual episodic memory and its context: autobiographical memory. *Philos Trans R Soc Lond B Biol Sci* 356(1413):1413. <https://doi.org/10.1098/rstb.2001.0940>
- Conway MA (2005) Memory and the self. *J Mem Lang* 53(4):594–628. <https://doi.org/10.1016/j.jml.2005.08.005>
- Diemer J, Alpers GW, Peperkorff HM, Shibani Y, Mühlberger A (2015) The impact of perception and presence on emotional reactions: a review of research in virtual reality. *Front Psychol*. <https://doi.org/10.3389/fpsyg.2015.00026>
- Diemer J, Mühlberger A, Pauli P, Zwanzger P (2014) Virtual reality exposure in anxiety disorders: impact on psychophysiological reactivity. *World J Biol Psychiatry* 15(6):6. <https://doi.org/10.3109/15622975.2014.892632>
- Dolcos F, Cabeza R (2002) Event-related potentials of emotional memory: encoding pleasant, unpleasant, and neutral pictures. *Cogn Affect Behav Neurosci* 2(3):252–263. <https://doi.org/10.3758/CABN.2.3.252>
- Erk S, Kiefer M, Grothe J, Wunderlich AP, Spitzer M, Walter H (2003) Emotional context modulates subsequent memory effect. *Neuroimage* 18(2):2. [https://doi.org/10.1016/S1053-8119\(02\)00015-0](https://doi.org/10.1016/S1053-8119(02)00015-0)
- Estupiñán S, Rebelo F, Noriega P, Ferreira C, Duarte E (2014) Can Virtual Reality Increase Emotional Responses (Arousal and Valence)? A Pilot Study. In: Marcus A (ed) *Design, User Experience, and Usability. User Experience Design for Diverse Interaction Platforms and Environments*. Springer International Publishing, Berlin Heidelberg, pp 541–549
- Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (Third). Sage. <https://socialsciences.mcmaster.ca/jfox/Books/Companion/>
- Gao C, Ren J, Sakaki M, Jia X (2024) Memory enhancement for emotional words is attributed to both valence and arousal. *Acta Psychol* 246:104249. <https://doi.org/10.1016/j.actpsy.2024.104249>
- Godden D, Baddeley A (1975) Context-dependent memory in two natural environments: on land and underwater. *Br J Psychol* 66(3):3. <https://doi.org/10.1111/j.2044-8295.1975.tb01468.x>
- Godden D, Baddeley A (1980) When does context influence recognition memory? *Br J Psychol* 71(1):99–104
- Gonçalves R, Pedrozo AL, Coutinho ESF, Figueira I, Ventura P (2012) Efficacy of virtual reality exposure therapy in the treatment of PTSD: a systematic review. *PLoS ONE* 7(12):e48469. <https://doi.org/10.1371/journal.pone.0048469>
- Green DM, Swets JA (1966) *Signal detection theory and psychophysics*. Wiley, New York
- Gupta, K., Aeron, S., Zhao, J., Abouelenin, J., Pai, Y. S., & Billinghamurst, M. (2025). Exploring emotional memory encoding and recall in virtual reality. *Proceedings of the extended abstracts of the chi conference on human factors in computing systems*, 1–7. <https://doi.org/10.1145/3706599.3719896>
- Gupta, K., Chan, S. W. T., Pai, Y. S., Strachan, N., Su, J., Sumich, A., Nanayakkara, S., & Billinghamurst, M. (2022). Total VREcall: using biosignals to recognize emotional autobiographical memory in virtual reality. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 6(2), 55:1–55:21. <https://doi.org/10.1145/3534615>
- Hamann S (2001) Cognitive and neural mechanisms of emotional memory. *Trends Cogn Sci* 5(9):394–400. [https://doi.org/10.1016/S1364-6613\(00\)01707-1](https://doi.org/10.1016/S1364-6613(00)01707-1)
- Harris CR, Pashler H (2005) Enhanced memory for negatively emotionally charged pictures without selective rumination. *Emotion* 5(2):191–199. <https://doi.org/10.1037/1528-3542.5.2.191>
- Harvey AG, Lee J, Smith RL, Gumport NB, Hollon SD, Rabe-Hesketh S, Hein K, Dolsen MR, Hamen K, Kanady JC, Thompson MA, Abrons D (2016) Improving outcome for mental disorders by enhancing memory for treatment. *Behav Res Ther* 81:35–46. <https://doi.org/10.1016/j.brat.2016.03.007>
- Hautus MJ (1995) Corrections for extreme proportions and their biasing effects on estimated values of d' . *Behav Res Methods Instrum Comput* 27(1):46–51. <https://doi.org/10.3758/BF03203619>

- Hothorn T, Bretz F, Westfall P (2008) Simultaneous inference in general parametric models. *Biom J* 50(3):3. <https://doi.org/10.1002/bimj.200810425>
- Johnsdorf M, Kisker J, Gruber T, Schöne B (2023a) Comparing encoding mechanisms in realistic virtual reality and conventional 2D laboratory settings: Event-related potentials in a repetition suppression paradigm. *Front Psychol*. <https://doi.org/10.3389/fpsyg.2023.1051938>
- Johnsdorf M, Pham KA, Schmidt T, Truong VL, Wohnig A, Kisker J, Gruber T, Schöne B (2023b) New is not always better: virtual reality does not necessarily enhance mnemonic processing. *Front Psychol*. <https://doi.org/10.3389/fpsyg.2023.1089725>
- Johnsdorf M, Kisker J, Sagehorn M, Gruber T, Schöne B (2025) Memory performance as a function of modality and context: revealing the cognitive mechanisms underlying enhanced memory in virtual reality. *Virtual Reality* 29(3):128. <https://doi.org/10.1007/s10055-025-01205-1>
- Kensinger EA (2009) Remembering the details: effects of emotion. *Emot Rev* 1(2):99–113. <https://doi.org/10.1177/1754073908100432>
- Kim H, Park JY, Kim KK (2018) Spatial learning and memory using a radial arm maze with a head-mounted display. *Psychiatry Investig* 15(10):935–944. <https://doi.org/10.30773/pi.2018.06.28.3>
- Kisker J, Gruber T, Schöne B (2019a) Behavioral realism and lifelike psychophysiological responses in virtual reality by the example of a height exposure. *Psychol Res*. <https://doi.org/10.1007/s00426-019-01244-9>
- Kisker J, Gruber T, Schöne B (2019b) Experiences in virtual reality entail different processes of retrieval as opposed to conventional laboratory settings: a study on human memory. *Curr Psychol*. <https://doi.org/10.1007/s12144-019-00257-2>
- Kisker J, Gruber T, Schöne B (2020) Virtual reality experiences promote autobiographical retrieval mechanisms: electrophysiological correlates of laboratory and virtual experiences. *Psychol Res*. <https://doi.org/10.1007/s00426-020-01417-x>
- Kothgassner OD, Goreis A, Kafka JX, Van Eickels RL, Plener PL, Felnhofner A (2019) Virtual reality exposure therapy for posttraumatic stress disorder (PTSD): a meta-analysis. *Eur J Psychotraumatol* 10(1):1654782. <https://doi.org/10.1080/20008198.2019.1654782>
- Kredlow MA, Eichenbaum H, Otto MW (2018) Memory creation and modification: enhancing the treatment of psychological disorders. *Am Psychol* 73(3):269–285. <https://doi.org/10.1037/amp0000185>
- Krokos E, Plaisant C, Varshney A (2019) Virtual memory palaces: immersion aids recall. *Virtual Reality* 23(1):1–15. <https://doi.org/10.1007/s10055-018-0346-3>
- Kuhne C, Kecelioglu ED, Maltby S, Hood RJ, Knott B, Ditton E, Walker FR, Kluge MG (2023) Direct comparison of virtual reality and 2D delivery on sense of presence, emotional and physiological outcome measures. *Front Virtual Real* 4:52. <https://doi.org/10.3389/frvir.2023.121100>
- Kulke L, Pasqualetto L (2024) Learning from 360-degree videos on a VR headset compared to 2D videos on a monitor. *Cogent Educ* 11(1):2355386. <https://doi.org/10.1080/2331186X.2024.2355386>
- Lamers MH, Lanen M (2021) Changing between virtual reality and real-world adversely affects memory recall accuracy. *Front Virtual Reality*. <https://doi.org/10.3389/frvir.2021.602087>
- Lane RD, Ryan L, Nadel L, Greenberg L (2015) Memory reconsolidation, emotional arousal, and the process of change in psychotherapy: new insights from brain science. *Behav Brain Sci* 38:e1. <https://doi.org/10.1017/S0140525X14000041>
- Lemay DJ, Bazelaïs P, Doleck T (2021) Transition to online learning during the COVID-19 pandemic. *Computers Human Behav Rep* 4:100130. <https://doi.org/10.1016/j.chbr.2021.100130>
- Makowski D, Sperduti M, Nicolas S, Piolino P (2017) “Being there” and remembering it: presence improves memory encoding. *Conscious Cogn* 53:194–202. <https://doi.org/10.1016/j.concog.2017.06.015>
- Mania K, Troscianko T, Hawkes R, Chalmers A (2003) Fidelity metrics for virtual environment simulations based on spatial memory awareness states. *Presence Teleoperators Virtual Environ* 12(3):296–310. <https://doi.org/10.1162/105474603765879549>
- Martarelli CS, Chiquet S, Ertl M (2023) Keeping track of reality: embedding visual memory in natural behaviour. *Memory* 31(10):1295–1305. <https://doi.org/10.1080/09658211.2023.2260148>
- Monaro M, Mazza C, Colasanti M, Colicino E, Bosco F, Ricci E, Biondi S, Rossi M, Roma P (2024) Testing memory of a VR environment: comparison with the real environment and 2D pictures. *Virtual Reality* 28(2):100. <https://doi.org/10.1007/s10055-024-00999-w>
- Montana JJ, Matamala-Gomez M, Maisto M, Mavrodiev PA, Cavallera CM, Diana B, Mantovani F, Realdon O (2020) The benefits of emotion regulation interventions in virtual reality for the improvement of wellbeing in adults and older adults: a systematic review. *J Clin Med* 9(2):500. <https://doi.org/10.3390/jcm9020500>
- Murnane K, Phelps MP (1993) A global activation approach to the effect of changes in environmental context on recognition. *J Exp Psychol Learn Mem Cogn* 19(4):882–894. <https://doi.org/10.1037/0278-7393.19.4.882>
- Murnane K, Phelps MP, Malmberg K (1999) Context-dependent recognition memory: the ICE theory. *J Exp Psychol Gen* 128(4):4. <https://doi.org/10.1037/0096-3445.128.4.403>
- Murre JMJ (2021) The Godden and Baddeley (1975) experiment on context-dependent memory on land and underwater: a replication. *R Soc Open Sci* 8(11):200724. <https://doi.org/10.1098/rso.200724>
- Nevin JA (1969) Signal detection theory and operant behavior: a review of David M. Green and John A. Swets’ Signal detection theory and psychophysics I. *J Exp Anal Behav* 12(3):475–480. <https://doi.org/10.1901/jeab.1969.12-475>
- Parsons TD (2015) Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences. *Front Human Neurosci* 9:52. <https://doi.org/10.3389/fnhu.2015.00660>
- Pavic K, Chaby L, Gricourt T, Vergilino-Perez D (2023) Feeling virtually present makes me happier: the influence of immersion, sense of presence, and video contents on positive emotion induction. *Cyberpsychol Behav Soc Netw* 26(4):238–245. <https://doi.org/10.1089/cyber.2022.0245>
- Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team. (2019). *nlme: Linear and Nonlinear Mixed Effects Models* (Version 3.1–140) [Computer software]. <https://CRAN.R-project.org/package=nlme>
- R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Riva G, Mantovani F, Capideville CS, Preziosa A, Morganti F, Villani D, Gaggioli A, Botella C, Alcañiz M (2007) Affective interactions using virtual reality: the link between presence and emotions. *Cyberpsychol Behav* 10(1):45–56. <https://doi.org/10.1089/cpb.2006.9993>
- Roa SMC, Rodríguez BJ (2024) Neurofeedback-driven emotional regulation training in a virtual reality environment: a machine learning approach using OpenBCI. *Int J Psychol Res* 17(2):2. <https://doi.org/10.21500/20112084.7467>
- Rocabado F, Duñabeitia JA (2024) Clouded judgments? The role of virtual weather in word valence evaluations. *Cogn Emot*. <https://doi.org/10.1080/02699931.2024.2354488>

- Roediger, H. L. I., & Marsh, E. J. (2003). Episodic and autobiographical memory. In *Handbook of psychology: Experimental psychology, Vol. 4*. (pp. 475–497). John Wiley & Sons, Inc.
- RStudio Team. (2020). *RStudio: integrated development environment for R*. RStudio, PBC. <http://www.rstudio.com/>
- Rubo M, Messerli N, Munsch S (2021) The human source memory system struggles to distinguish virtual reality and reality. *Computers Human Behav Rep* 4:100111. <https://doi.org/10.1016/j.chbr.2021.100111>
- Sarfan LD, Zieve G, Gumpert NB, Xiong M, Harvey AG (2023) Optimizing outcomes, mechanisms, and recall of cognitive therapy for depression: dose of constructive memory support strategies. *Behav Res Ther* 166:104325. <https://doi.org/10.1016/j.brat.2023.104325>
- Schöne B, Kisker J, Lange L, Gruber T, Sylvester S, Osinsky R (2023) The reality of virtual reality. *Front Psychol* 14:52. <https://doi.org/10.3389/fpsyg.2023.1093014>
- Schöne B, Kisker J, Sylvester RS, Gruber T (2021) Library for universal virtual reality experiments (luVRe): a standardized immersive 3D/360° picture and video database for VR based research. *Curr Psychol*. <https://doi.org/10.1007/s12144-021-01841-1>
- Schöne B, Wessels M, Gruber T (2019a) Experiences in virtual reality: a window to autobiographical memory. *Curr Psychol* 38(3):3. <https://doi.org/10.1007/s12144-017-9648-y>
- Schöne B, Wessels M, Gruber T (2019b) Experiences in virtual reality: a window to autobiographical memory. *Curr Psychol* 38(3):715–719. <https://doi.org/10.1007/s12144-017-9648-y>
- Schröder D, Wrona KJ, Müller F, Heinemann S, Fischer F, Dockweiler C (2023) Impact of virtual reality applications in the treatment of anxiety disorders: a systematic review and meta-analysis of randomized-controlled trials. *J Behav Ther Exp Psychiatry* 81:101893. <https://doi.org/10.1016/j.jbtep.2023.101893>
- Shahid S, Kelson J, Saliba A (2024) Effectiveness and user experience of virtual reality for social anxiety disorder: systematic review. *JMIR Ment Health* 11(1):e48916. <https://doi.org/10.2196/48916>
- Shin YS, Masis-Obando R, Keshavarzian N, Dáve R, Norman KA (2021) Context-dependent memory effects in two immersive virtual reality environments: on Mars and underwater. *Psychon Bull Rev* 28(2):574–582. <https://doi.org/10.3758/s13423-020-01835-3>
- Smith SA (2019) Virtual reality in episodic memory research: a review. *Psychon Bull Rev*. <https://doi.org/10.3758/s13423-019-01605-w>
- Smith SM, Manzano I (2010) Video context-dependent recall. *Behav Res Methods* 42(1):292–301. <https://doi.org/10.3758/BRM.42.1.292>
- Smith SM, Vela E (2001) Environmental context-dependent memory: a review and meta-analysis. *Psychon Bull Rev* 8(2):2. <https://doi.org/10.3758/BF03196157>
- Somarathna R, Bednarz T, Mohammadi G (2023) Virtual reality for emotion elicitation – a review. *IEEE Trans Affect Comput* 14(4):2626–2645. <https://doi.org/10.1109/TAFFC.2022.3181053>
- Tulving E, Thomson DM (1973) Encoding specificity and retrieval processes in episodic memory. *Psychol Rev* 80(5):352–373. <https://doi.org/10.1037/h0020071>
- van Helvoort D, Stobbe E, Benning R, Otgaar H, van de Ven V (2020) Physical exploration of a virtual reality environment: effects on spatiotemporal associative recognition of episodic memory. *Mem Cognit* 48(5):691–703. <https://doi.org/10.3758/s13421-020-0102-4-6>
- Wälti MJ, Woolley DG, Wenderoth N (2019) Reinstating verbal memories with virtual contexts: myth or reality? *PLoS ONE* 14(3):e0214540. <https://doi.org/10.1371/journal.pone.0214540>
- Wang R, Yu R, Tian Y, Wu H (2022) Individual variation in the neurophysiological representation of negative emotions in virtual reality is shaped by sociability. *Neuroimage* 263:119596. <https://doi.org/10.1016/j.neuroimage.2022.119596>
- Wardell V, Palombo DJ (2024) Stability and malleability of emotional autobiographical memories. *Nat Rev Psychol* 3(6):393–406. <https://doi.org/10.1038/s44159-024-00312-1>
- Watkins MJ, Ho E, Tulving E (1976) Context effects in recognition memory for faces. *J Verbal Learn Verbal Behav* 15(5):505–517. [https://doi.org/10.1016/0022-5371\(76\)90045-1](https://doi.org/10.1016/0022-5371(76)90045-1)
- Wessels M, Valuch C, Schacht A, Kulke L (2019a) Differences in incidental memory of virtual reality compared to 2D videos of emotional scenes. <https://doi.org/10.17605/OSF.IO/FUJPR>
- Wessels M, Valuch C, Schacht A, Kulke L (2019b) Retrieving incidental memories in Virtual Reality. <https://doi.org/10.17605/OSF.IO/Z76SJ>
- Williams SE, Ford JH, Kensinger EA (2022) The power of negative and positive episodic memories. *Cogn Affect Behav Neurosci* 22(5):869–903. <https://doi.org/10.3758/s13415-022-01013-z>
- Wong KP, Lai CYY, Qin J (2023) Systematic review and meta-analysis of randomised controlled trials for evaluating the effectiveness of virtual reality therapy for social anxiety disorder. *J Affect Disord* 333:353–364. <https://doi.org/10.1016/j.jad.2023.04.043>
- Wu J, Sun Y, Zhang G, Zhou Z, Ren Z (2021) Virtual reality-assisted cognitive behavioral therapy for anxiety disorders: a systematic review and meta-analysis. *Front Psych*. <https://doi.org/10.3389/fpsyg.2021.575094>
- M Yu Y Bai Y Li 2023 Emo-regulator: An emotion-regulation training system fusing virtual reality and EEG-based neurofeedback. Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society Annual International Conference 2023 1 4 <https://doi.org/10.1109/EMBC40787.2023.10340975>
- Zaki J, Ochsner K (2009) The need for a cognitive neuroscience of naturalistic social cognition. *Ann N Y Acad Sci* 1167(1):16–30. <https://doi.org/10.1111/j.1749-6632.2009.04601.x>

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