Applied Cognitive Psychology

WILEY

RESEARCH ARTICLE OPEN ACCESS

Virtual Reality Provides an Eyewitness Experience That Is Similar to Real Life

Andrew D. Green[1,2] | Andrew Clark[1] | Joanne Rechdan[3] | Andy Guppy[1]

[1]School of Psychology, University of Bedfordshire, Luton, UK | [2]School of Psychology, Arden University, Coventry, UK | [3]Department of Psychology, Bournemouth University, Poole, UK

Correspondence: Andrew D. Green (agreen@arden.ac.uk)

Received: 25 February 2025 | **Revised:** 8 July 2025 | **Accepted:** 13 July 2025

Funding: The authors received no specific funding for this work.

Keywords: eyewitness memory | presence | virtual reality

ABSTRACT

Traditional methods used for presenting to-be-remembered events in eyewitness memory research are often criticized for lacking scientific rigor. Videos lack ecological validity, and though staged live events are realistic, they lack experimental control. Virtual reality (VR) has been proposed as a promising alternative, offering immersive realism in a controlled environment. In this study, 141 participants viewed an event either live, on video, or in VR. Presence, emotional experience, heart rate, and recall were compared across groups, and it was seen that the VR experience was highly similar to the live-event group. The video group reported significantly lower presence, ecological validity, and heart rate changes compared to the VR group. These findings suggest that VR can offer a highly realistic witness experience while maintaining experimental control, making it a valuable tool for eyewitness memory research.

1 | Introduction

Eyewitness memory reports are a critical form of evidence in the criminal justice system. Decades of research in the area suggest that memory is highly susceptible to error and manipulation and is a leading cause of wrongful conviction (Innocence Project 2025). A substantial body of research has demonstrated that eyewitness memory is not always accurate and that memory for past experiences can be manipulated with ease (see Loftus 2005 for a review). Commonly, in this research, participants are exposed to a to-be-remembered stimulus event (such as a crime) and their memory is subsequently assessed, perhaps following a delay or manipulation. The event is often displayed as a video (Foster et al. 2012) or as a live event (Rubínová et al. 2021), though other mediums are used such as slides (Loftus et al. 1978) or clips from existing media (LaPaglia et al. 2014). However, these methods are frequently criticized for lacking scientific rigor.

Laboratory studies, which often use video stimuli, are criticised for lacking ecological validity. Videos may offer a perfect view of events, which may not be the case in real life (Ihlebæk et al. 2003), and are viewed in the safety of a lab environment, under the supervision of a researcher. As such, it can be challenging to elicit realistic physiological responses from participants, which have been shown to impact eyewitness memory (Deffenbacher et al. 2004). Stress during encoding has been shown to negatively affect eyewitness' free recall (Metcalfe et al. 2019) and cued recall (Stanny and Johnson 2000). Experiencing real world threat can change attentional demands of witnesses, as they may be assessing the situation (Chae 2010) rather than trying to memorise details. Watching a video of a crime does not provide a realistic witness experience, and thus the results are difficult to generalise to the real world. However, it is a popular method of displaying events, as videos are easy to create and manipulate, and are highly reliable. Laboratory studies allow researchers high levels of internal validity, at the cost of ecological validity.

An alternative approach that offers high levels of ecological validity is live events. A to-be-remembered event is portrayed in person by actors and is often intended to be believed as real by participants. These events can be staged in front of large groups of participants, such as students in a lecture (Buckhout et al. 1975; Flin et al. 1992). If participants believe the event to be real, this should lead to responses that are congruent with seeing an actual crime occur. This approach can be less reliable as actors may make errors or vary behavior between performances, and participants may try to intervene. Conducting live events with large groups reduces repetitions and thus improves reliability; however, this means that any mistakes that do occur will impact many participants (Wells and Penrod 2011). Using smaller groups or presenting events to individuals can limit the impact of mistakes, but the larger number of repetitions increases the chances of mistakes happening in the first place.

As live events are often designed so that participants believe a real crime is being committed, there are substantial ethical concerns, and it may not be possible to present distressing or violent crimes. Some researchers address this issue by presenting forensically relevant, non-crime events. For example, Bates et al. (1999) compared the recall of child eyewitnesses who witnessed an event live or on video. The to-be-remembered event involved a librarian searching for a lost bag in the school library, finding more accurate recall following a live event compared to a video. Due to the need for less distressing stimuli in live events, perhaps videos are the more ecologically valid option for certain research questions, such as those involving violent offenses or weapons, as they more easily accommodate these types of crime. Additionally, it may be difficult to relate the findings of non-crime research to the experience of witnessing crimes in the real world. However, in general, live events can allow for high levels of ecological validity but decrease reliability and internal validity.

Ihlebæk et al. (2003) staged a crime event in front of live eyewitnesses, recorded this, and presented the video to a video condition. When comparing the memory of the two groups, they found that participants who saw a video stimulus recalled more correct information than those who experienced the live event. They proposed that participants in video studies see the event take place in ideal and unrealistic circumstances and that these studies may overestimate memory performance. While videos can provide participants a uniform viewing experience, the perspectives of live-event participants may vary. Although this might replicate differences between eyewitnesses to real crimes, it also introduces confounding variables such as viewing distance, which can impact memory (Semmler et al. 2018). The contrasting findings of Bates et al. (1999) and Ihlebæk et al. (2003), as well as the issues they highlight, demonstrate the need to further compare these methods and identify the best way to present stimuli in eyewitness memory experiments.

Kothgassner and Felnhofer (2020) discuss the compromise many researchers have to make between ecological validity and experimental control, and suggest virtual reality (VR) as a potential solution for the issue. VR refers to technology that is used to surround a user's senses with a computer-generated environment (Slater 2018). The virtual environments can be entirely computer-generated, like a video game, or they can be 360° videos recorded in the real world. These experiences surround the user in all directions, allowing them to turn their head and explore the environment as if they were actually there. In addition to providing a more realistic visual experience than a video, VR can elicit a sense of presence. This term describes the feeling of "being there" (Reeves 1991) within a virtual environment. Users experience a cognitive illusion of being located within the virtual world rather than the physical one, temporarily forgetting that the experience is mediated by technology (Lombard and Ditton 1997). Though users know at some level that they are interacting with a virtual environment, they may respond to that world as if it were real (Slater 2018). Research has shown that VR can elicit realistic physiological, behavioral (Kisker et al. 2021; van Dammen et al. 2022) and emotional responses (Felnhofer et al. 2015). If VR can provoke real-world responses, then this is an indicator of ecological validity (Kothgassner and Felnhofer 2020). Additionally, the use of 360-videos allows researchers to maintain the reliability and control of a traditional laboratory video study, suggesting that VR is a research tool that is worth investigating for eyewitness research.

In a scoping review of studies examining the differences between VR, videos, and the realorld, Hepperle and Wölfel (2023) found that VR headsets provide experiences that are more similar to the real world than to videos and that there are more similarities between VR and the real world than differences. For example, they reviewed 57 perception studies which compared VR to video or a live event, including topics such as emotion, engagement, learning, realism, and presence. Generally, they found that there was no difference between VR and live events in these studies, while VR was significantly different from videos in almost all instances. VR has been shown to elicit similar levels of emotion and presence to live events (Chirico and Gaggioli 2019), suggesting a high level of ecological validity. Ceccato et al. (2024) exposed participants to an office either as 2D pictures on a screen, in VR, or in real life, and it was found that the real-life group had the best memory performance, and there was no difference between the 2D and VR groups. Though this was a recall task, the experiment included no social actors or narrative events; therefore, the findings may not generalize to eyewitness memory. Kim et al. (2014), Kloft et al. (2020) and Glomb et al. (2024) chose to display to-be-remembered events to participants in eyewitness memory experiments, as VR can provide a realistic experience that can induce a sense of presence. These studies did not explore VR or presence as variables of interest; rather, the medium was used to explore expectation and arousal, substance use and sleep restriction, and cognitive styles, respectively. Nyman et al. (2020) compared video stimuli to a 360° VR video in a recognition (rather than recall) study, where they found that the video group was more accurate. This is in line with the view of Ihlebæk et al. (2003) that studies involving video stimuli may overestimate performance.

Attention allocation, physiological arousal, and emotional experience are connected factors that can influence memory (Wulff and Thomas 2021). Witmer and Singer (1998) suggest that presence in a virtual environment is directly related to how much attention is directed to the virtual and real world. This is supported by Kober and Neuper (2012) who have demonstrated that increasing presence also increases activity in brain regions associated with allocating attention. This implies that high levels of presence correlate with high levels of attention, which in turn has implications for memory encoding (Chun and Turk-Browne 2007). Presence has been shown to correlate with levels of emotion experienced (Dibbets and Schulte-Ostermann 2015), and there is evidence that people are more likely to recall emotional than neutral events (Kensinger and Schacter 2006). This emotional enhancement of memory occurs as emotional stimuli (positive or negative) attract more attention than neutral, and more encoding resources are applied, enhancing storage and recall of related information (Kensinger and Corkin 2004). This attention-mediation hypothesis is mostly based on non-VR studies (Cadet and Chainay 2020); however, there has been some research investigating the link between VR, emotion, and memory. Schöne et al. (2019) had participants experience what they considered a highly emotional motorcycle ride in VR and on a desktop computer. Recall was better in the VR condition, and mood positively correlated with recall.

Our previous work (Green et al. 2025) sought to directly compare eyewitness memory, arousal, and emotional experience between a video and VR group. Across two studies, viewing a to-be-remembered crime event in VR elicited a statistically significantly higher sense of presence and involvement in the scene. In the first study, participants in the VR group had a significantly higher heartrate during that event compared to the video group. In the second study, participants in the VR group reported a more emotional experience than the video group. These results suggest that participants in the VR condition may have had a more realistic and ecologically valid eyewitness experience. There was no difference in recall accuracy between the groups in either experiment. These results are consistent with the findings of Glomb et al. (2024) who conducted a similar study. Both Green et al. (2025) and Glomb et al. (2024) suggest that though VR may be seen as more realistic than videos in their studies, a live event is needed for comparison.

The current study aims to build on those studies outlined by Green et al. (2025). Participants were presented a to-be-remembered event live, on video, or in VR. Memory, sense of presence, emotional experience, and arousal were compared to establish if VR is a suitable alternative method of stimulus presentation that provides an ecologically valid experience. To date, no eyewitness memory research has compared these three groups. In line with Ihlebæk et al. (2003) it was hypothesized that the video group would have significantly better recall of events compared to the VR or live-event group [H1], and that there would be no significant difference in recall between the VR and live-event groups [H2]. It was also expected that the live-event group would have a significantly greater sense of presence than the other two groups [H3], while the VR group would have a greater sense of presence than the video group [H4]. As it is anticipated that the video group will be the least realistic, it is hypothesized that they would have a less emotional experience than the other groups [H5] and that there would be no significant difference between the VR and live-event groups [H6]. As physiological arousal can be directly linked to the level of presence experienced, it was expected that presence would have a significant positive correlation with heart rate (HR) [H7], and that the video group would experience a significantly lower HR than the VR group [H8].

All hypotheses were preregistered on the open science framework (https://osf.io/e39nz) prior to data collection.

2 | Method

2.1 | Design and Participants

The study was approved by the Psychology ethics committee of the University of Bedfordshire. A power analysis using effect sizes from Green et al. (2025) suggested a minimum of 41 participants per group (power = 0.80, F = 0.29, alpha = 0.05) to detect significant differences in presence. Fifty-five participants were present and took part in the initial live condition, with additional data collected beyond the power analysis to avoid falling short of the desired sample size if participant data had to be removed. Eight participants in the live-event group provided incomplete data (only signing the consent form or providing basic demographic information) and were removed from analysis. Participant numbers for the video and VR group were then matched to the live-event group, with 47 taking part in each condition. The authors acknowledge that this slightly exceeds the preregistered sample size; however, the target sample size for the VR and video groups (based on the live group attendance) was set prior to data collection in those groups.

Overall, 141 participants (98 female, 37 male, 6 did not disclose) aged 18–63 (M = 31.16, SD = 11.16) took part in the study. The live-event group consisted of a cohort of third-year undergraduate psychology students from the University of Bedfordshire who viewed the to-be-remembered event in a scheduled lecture. The video and VR groups were invited to take part in a memory study through opportunity sampling. First-and second-year psychology undergraduate students were recruited through the SONA participant pool and received credits required for their research methods unit. All participants were offered the opportunity to enter a prize draw for an Amazon voucher. The total sample consisted of psychology students (75.89%) and staff (6.38%), students and staff from other disciplines within the university (4.26% and 8.51% respectively), and members of the public (4.96%).

Participants in the live-event group were not randomly allocated to the group, as it was a surprise event that took place during a scheduled lecture. However, participants in the video and VR groups were randomly allocated using Qualtrics.

2.2 | Materials

2.2.1 | Stimulus Event and Recording

Two actors entered a university lecture and had an argument with the lecturer regarding their grades in front of the student cohort. After approximately 2 min, the actors left the room, pushing items to the floor on the way out. Although the disruptive individuals were actors, and the scene was staged, the intention was that the student audience present in the room would perceive it as real. The to-be-remembered event was filmed using an Insta360 OneX2, 360° camera (Video version: https://osf.io/27qja; 360 version: https://osf.io/txdmq). Awareness of the camera during the live event could have made participants suspect that the events taking place were not real. To reduce this suspicion, the camera was present at the lecture from the beginning of the semester and was occluded from participants line of sight. This video was then used as the stimuli for the video and VR groups. All participants saw the exact same stimuli, with the only difference being the medium of viewing (real life, video screen or VR HMD).

A HTC Vive Pro 2 virtual reality head-mounted display was used to display the stimuli to participants in the VR group, and a 21″ computer monitor was used to display to the video group. A Powerlab 26T and pulse transducer from ADInstruments was used to record HR, and this was analysed using Lab Chart 7.

2.2.2 | Scales

Experiences in VR can cause visually induced motion sickness (VIMS), so the Simulator Sickness Questionnaire (SSQ; Kennedy et al. 1993) was used before and after viewing the video or VR stimulus to measure well-being and likelihood of experiencing VIMS, with high scores leading to the termination of the experiment. The experiments described by Green et al. (2025) used the IPQ (Schubert et al. 2001) to measure presence. Though this scale is reliable and consists of subfactors relevant to this research, it is applicable only to digital stimuli as it specifically asks participants about the "virtual world". Instead, the Independent Television Commission Sense of Presence Inventory (ITC-SOPI; Lessiter et al. 2001) was used as it was specifically designed as a cross-media presence scale that can be used beyond VR and can discriminate between real and virtual worlds (Busch et al. 2014; Nisenfeld 2003). This is a 38-item scale that assesses spatial presence, engagement, ecological validity, and negative effects of stimuli. Spatial presence refers to the sense of physical presence within the scene, engagement refers to psychological involvement, and ecological validity refers to how real or life-like the experience was. Negative effects describe physiological symptoms such as nausea and eyestrain. The subscales have been shown by the authors to be reliable ($\alpha = 0.76 - \alpha = 0.94$), sensitive to different media, and valid when compared to corroborative measures. Participants respond to statements on a 5-point Likert scale running from strongly disagree to strongly agree.

The Affective Slider (AS; Betella and Verschure 2016) was used to measure overall valence of the scene and self-reported arousal. This scale, consisting of two illustrated questions, uses a slider that runs from 0 (negative emotion or low arousal) to 1 (positive emotion or high arousal). This scale was proposed as a simpler alternative to the Self-Assessment Manikin (Bradley and Lang 1994) to which it strongly correlates. The Positive and Negative Affect Schedule (PANAS; Watson et al. 1988) was also used to measure the emotional experience of participants. This scale has 20 self-report items such as guilt, fear, and enthusiasm, as well as providing overall positive (PA) and negative (NA) affect scores. The PANAS has high reliability, with Cronbach's alphas ranging from 0.86–0.90 (positive affect) and 0.84–0.87 (negative affect).

In the video and VR group, all measures were administered using the Qualtrics platform. In the live-event group, this was done on paper and entered into Qualtrics later by a research assistant.

2.2.3 | Recall Tasks

To assess participants' memory accuracy, free and cued recall tasks were completed. In the free recall task, participants were asked to write as much information about the event as they could possibly recall, and were encouraged to provide specific details where possible. The cued recall task asked participants specific questions about the event. Usually, cued recall questions would be developed in a pilot study, where participants are shown the target stimulus and asked to recall events. This free recall would then be the basis of the cued questions (as described by Wilford et al. 2014). However, in the current study the video was recorded minutes before the live-event group would be asked to complete free and cued recall tasks, so a pilot was not possible. Instead, questions were developed from those used by Green et al. (2025). For example, as there were several questions that related to the appearance of the perpetrator, a question was developed in line with that theme: "What colour was the perpetrator's hair?" Questions were avoided that the live-event group might have an advantage in answering (e.g., "Where did the event take place?", as they were still in the lecture hall when answering the cued recall questions). In total, 12 cued recall questions were developed.

2.3 | Procedure

2.3.1 | Live-Event Group

During a final year psychology lecture, a short event occurred which served as the to-be-remembered stimulus for all three groups, as described in the materials. The lecture then continued for 10 min before it was revealed to be an experiment. Students in the room were then given the opportunity to volunteer to take part in the study and were provided a paper questionnaire to complete by hand. Participants were asked to provide demographic information and complete the ITC-SOPI, PANAS, and AS. Following these scales, they were asked to complete a free and cued recall task to assess their memory of the event. They were also offered the opportunity to ask questions and opt in to a prize draw. They were then fully debriefed and thanked for their time. HR recording was not possible in the live-event group due to the number of simultaneous participants and the surprise nature of the scenario.

2.3.2 | Video and VR Groups

The video and VR conditions took place one participant at a time, in a research laboratory at the University of Bedfordshire. Participants in these groups were asked to complete the SSQ to assess their current wellbeing, with a high score leading to termination of the experiment. However, no participants were excluded on this basis. They were then randomly allocated to either the video or VR group by Qualtrics and provided demographic information. Baseline HR was recorded for 1 min, and participants then watched the stimulus video either on screen or in virtual reality while the HR recording continued. Following the stimulus video, participants completed the ITC-SOPI, AS, and PANAS.

Participants then completed a 10-min distractor task to simulate the 10-min delay in the live-event group and were then asked to complete the free and cued recall tasks. Participants were again offered the opportunity to ask questions and opt into the prize draw, as well as being debriefed and thanked for their time.

3 | Results

Hypotheses and planned data analysis were pre-registered prior to data collection (https://osf.io/e39nz) following the initial pre-registered analysis, several exploratory analyses were also conducted and are reported separately.

3.1 | Pre-Registered Data Analysis

3.1.1 | Recall

Participants' free recall was coded and scored following the guidelines from Wright and Holliday (2007). The stimulus video was coded into idea units, encompassing any information that could possibly be correctly recalled. This coding process was completed by the researcher and a research assistant separately, to ensure that all possible items were included. In total, 440 items were identified: 161 person, 140 action, 34 object, and 105 surrounding.

Once participants' free recall were collected, they were coded into these idea units, compared to the original list, and classed as correct, incorrect or confabulation. For example, in part of the scene, a woman knocks papers onto the floor. The participant response "the man pushed over the papers" would be marked as "the man (incorrect/person) pushed over (correct/action) the papers (correct/item)". Each participant recall was scored by the researcher and a research assistant who was not aware of the specific hypotheses of the study. From each account, the total number of correct, incorrect and confabulated items were recorded. An accuracy ratio was calculated as: $N_{correct}/(N_{correct} + N_{incorrect} + N_{confabulations})$. This value ranging from 0 to 1 represents the correct items as a proportion of a participant's entire account, with 0 being 0% accurate, and 1 representing 100% accuracy. Total number of items reported was also calculated by summing the correct, incorrect and confabulation scores. An intraclass correlation of 0.64 was found between the researcher and research assistant's accuracy ratios, 0.62 for total correct items and 0.58 for incorrect items suggesting moderate interrater reliability ($p < 0.001$). However, reliability was poor for confabulated items (0.20), likely due to the subjective difference between coding items as incorrect or confabulations. When incorrect and confabulated items were combined into a "non-correct" category, interrater reliability rose to 0.60, supporting this conclusion. Cued recall responses were scored as either correct, incorrect or a "not sure" response. A cued recall accuracy ratio was calculated as $N_{correct}/(N_{correct} + N_{incorrect})$. "Not sure" responses were not treated as incorrect, rather removed from the accuracy calculation. Like the free recall, the cued recall accuracy ratio can be directly interpreted as a percentage of accuracy. The descriptive statistics for these recall variables can be seen in Table 1.

Participants in the video and VR groups were 95% accurate in their free recall, and the live-event group was slightly higher at 96%. In the cued recall task, accuracy was 59% for the VR group, 63% for video, and 68% for the live-event group. The video group reported the most correct and incorrect items, as they reported the highest number of items in general. They also had the lowest number of "not sure" responses in the cued task. The VR group had the most confabulations and "not sure" responses, the lowest mean correct score in the cued task, and the lowest cued accuracy ratio. The live-event group had the lowest free recall correct, incorrect, and confabulation scores; thus, they also reported the least items. The live-event group reported the least incorrect items in the cued task and had the highest free and cued accuracy ratios.

Multivariate analyses were conducted to address hypotheses 1 and 2. A one-way MANOVA was conducted to compare the free recall scores of the three groups, including the outcome variables of total correct, incorrect, and confabulated items, and the accuracy ratios. Using Pillai's trace, there was a statistically significant difference between the groups, $V = 0.25$, $F(8, 272) = 4.90$, $p < 0.001$, $\eta p^2 = 0.13$. Univariate comparison of the free recall measures revealed that there was a significant difference between the groups in total correct items ($p < 0.001$, $\eta p^2 = 0.18$) and total incorrect items ($p = 0.01$, $\eta p^2 = 0.06$). Post hoc *t*-tests, employing the Bonferroni adjustment, showed that the live-event group recalled significantly fewer correct items (video, $p < 0.001$; VR, $p < 0.001$) and incorrect items (video, $p = 0.03$; VR, $p = 0.03$) than the other groups. Further analysis is reported in the exploratory analysis section, comparing the total number of items reported by each group. There was no significant difference between the video or VR groups in any of these variables ($p < 0.05$). There was no significant difference between the groups' total confabulations ($p = 0.05$, $\eta p^2 = 0.04$) or accuracy ratios ($p = 0.25$, $\eta p^2 = 0.02$). A one-way MANOVA was conducted to compare the cued recall scores across groups, including the outcome variables of total correct, incorrect, and "not sure" items, and the accuracy ratio. Two participants in the live-event group did not complete the cued recall task. Using Pillai's trace, there was no statistically significant difference between the groups, $V = 0.08$, $F(6, 270) = 1.90$, $p = 0.08$, $\eta p^2 = 0.04$. Overall, the findings suggest that though there was a difference in the quantity of freely recalled information, with the live-event group reporting less than the other groups, there was no difference in free or cued recall accuracy.

3.1.2 | Ecological Validity

Presence, HR, and emotional experience were investigated as indicators of ecological validity. To address hypotheses 3 and 4,

TABLE 1 | Means and standard deviation of recall scores across groups.

		Video	VR	Live
		Mean (SD) [95% CI]	**Mean (SD) [95% CI]**	**Mean (SD) [95% CI]**
Free recall	Total correct	32.68 (11.91)	31.17 (11.09)	21.62 (7.93)
		[29.28, 36.09]	[28.00, 34.34]	[19.35, 23.89]
	Total incorrect	1.43 (1.57)	1.40 (1.80)	0.60 (1.10)
		[0.98, 1.88]	[0.89, 1.92]	[0.28, 0.91]
	Total confabulation	0.36 (0.70)	0.68 (1.00)	0.32 (0.59)
		[0.16, 0.56]	[0.39, 0.97]	[0.15, 0.49]
	Accuracy ratio	0.95 (0.05)	0.95 (05)	0.96 (0.05)
		[0.93, 0.96]	[0.93, 0.96]	[0.95, 0.98]
	Total items	34.47 (12.78)	33.25 (12.76)	22.53 (8.29)
		[30.81, 38.12]	[29.61, 36.91]	[20.16, 24.90]
Cued recall	Total correct	6.13 (1.94)	5.74 (2.21)	6.56 (1.81)
		[5.56, 6.70]	[5.10, 6.39]	[6.01, 7.10]
	Total incorrect	3.60 (1.91)	3.85 (1.68)	3.09 (1.64)
		[3.04, 4.16]	[3.36, 4.35]	[2.60, 3.58]
	Total "not sure"	2.28 (1.64)	2.40 (1.75)	2.36 (1.65)
		[1.80, 2.76]	[1.89, 2.92]	[1.86, 2.85]
	Accuracy ratio	0.63 (0.18)	0.59 (0.18)	0.68 (0.16)
		[0.58, 0.69]	[0.53, 0.64]	[0.64, 0.73]

the four ITC-SOPI subfactors of spatial presence, engagement, ecological validity, and negative effects were compared across groups. As can be seen in Table 1, the ITC-SOPI scores for spatial presence and ecological validity are highest in the live-event group, which is expected as those who were physically in the room are reporting a greater feeling of "being there" and realness of the event, as well as the greatest level of negative impact from the event. The VR group scored highest for engagement, and the video group scored lowest on all presence factors. A one-way MANOVA was conducted to compare the ITC-SOPI presence scores across the groups, including the outcome variables of spatial presence, engagement, ecological validity, and negative effects. Using Pillai's trace, there was a statistically significant difference in the sense of presence between the groups, $V=0.430$, $F(8, 270)=9.24$, $p<0.001$, $\eta p^2=0.22$.

Univariate comparisons were conducted to investigate each factor in more depth. Post hoc *t*-tests, employing the Bonferroni adjustment, were then used to compare the groups on each factor. There was a statistically significant difference in spatial presence between the groups, $F(2, 137)=24.03$, $p<0.001$, $\eta p^2=0.26$. The video group experienced significantly less spatial presence than the VR and live-event groups ($p<0.001$). There was no significant difference between the VR and live-event groups ($p=0.98$). Participants in the live and VR groups had a similar sense of "being there" within a real physical space, which was much greater than the video group. There was a statistically significant difference in engagement between the groups, $F(2, 137)=4.70$, $p=0.01$, $\eta p^2=0.06$. The live-event group experienced a statistically similar level of engagement to the other groups (video, $p=0.81$; VR, $p=0.18$), however there was a significant difference between the VR and video groups ($p=0.009$), with the VR group reporting a greater sense of involvement in events than the video group. There was a statistically significant difference in ecological validity between the groups, $F(2, 137)=8.12$, $p<0.001$, $\eta p^2=0.20$. The live-event group scored significantly higher than the other groups (video, $p<0.001$; VR, $p=0.008$), and the VR group scored significantly higher than the video group ($p=0.02$). The live-event group rated the event highest in terms of realism, and the VR group found it more realistic than the video group. There was a statistically significant difference in negative effects between the groups, $F(2, 137)=7.64$, $p<0.001$, $\eta p^2=0.10$. The live-event group scored significantly higher than the video group ($p<0.001$) but similar to the VR group ($p=0.35$). There was no significant difference in score between the VR and video groups ($p=0.07$). The live-event group reported the highest level of physical discomfort during the scene, which was similar to the VR group, and higher than the video group. These results are illustrated in Figure 1.

To address hypotheses 5 and 6, the AS arousal, AS valence, PANAS PA, and PANAS NA were compared across groups. As displayed in Table 2, the video group had the lowest positive affect. The VR group scored highly on AS arousal, rated the valence the most positive, and had the lowest negative affect score on the PANAS. The live-event group had high scores for both positive and negative affect and the lowest self-report arousal. A one-way MANOVA was conducted to compare the emotion scores across the groups, including the outcome variables of AS arousal, AS valence, PANAS positive affect, and PANAS negative affect. Using Pillai's trace, there was a statistically significant difference in emotional experience between the groups, $V=0.162$, $F(8, 260)=9.24$, $p=0.004$, $\eta p^2=0.08$.

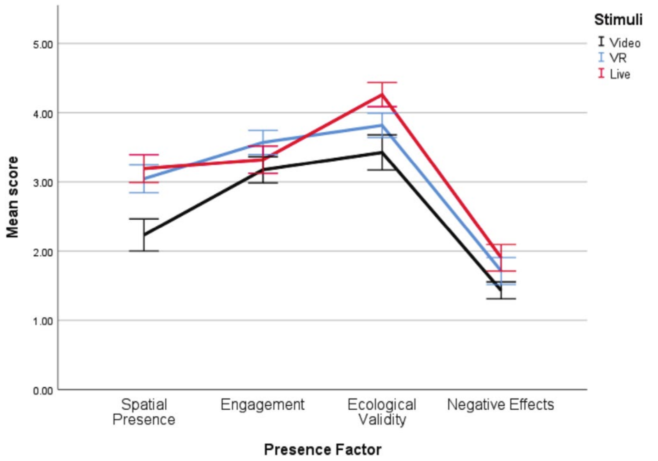

FIGURE 1 | Mean ITC-SOPI presence scores across conditions. Error bars represent 95% confidence intervals.

There was no statistically significant difference in the AS arousal [$F(2, 132)=2.90$, $p=0.06$, $\eta p^2=0.04$], AS valence [$F(2, 132)=2.54$, $p=0.08$, $\eta p^2=0.04$] or PANAS positive affect [$F(2, 132)=1.27$, $p=0.28$, $\eta p^2=0.02$] scores between groups. However, there was a statistically significant difference between the PANAS negative affect scores, $F(2, 132)=5.34$, $p=0.006$, $\eta p^2=0.08$. Post hoc *t*-tests, employing the Bonferroni adjustment, showed that the participants in the live-event group reported a significantly greater level of negative affect than the other groups (video, $p=0.047$; VR, $p=0.006$). There was no statistically significant difference between the video and VR groups ($p>0.05$). The live-event group reported feeling more negative emotions during the event than the other groups, which were similar to each other. The differences in emotional experience are illustrated in Figure 2.

To address hypothesis 7, that HR would correlate with presence, a series of correlations were conducted between experimental HR and the ITC-SOPI measures of presence. All variables were normally distributed except for ecological validity ($W=0.95$, $p=0.003$) and negative effects ($W=0.88$, $p<0.001$), which were analyzed using Spearman's

TABLE 2 | Mean and SD for measures of ecological validity across groups.

	Video	VR	Live
	Mean (SD) [95% CI]	Mean (SD) [95% CI]	Mean (SD) [95% CI]
Spatial presence	2.23 (0.79)	3.04 (0.68)	3.23 (0.72)
	[2.01, 2.46]	[2.85, 3.24]	[3.03, 3.44]
Engagement	3.18 (0.65)	3.57 (0.59)	3.36 (0.70)
	[2.99, 3.36]	[3.40, 3.74]	[3.16, 3.56]
Ecological validity	3.43 (0.86)	3.82 (0.60)	4.28 (0.59)
	[3.18, 3.67]	[3.65, 3.99]	[4.11, 4.45]
Negative effects	1.43 (0.42)	1.71 (0.66)	1.90 (0.64)
	[1.31, 1.55]	[1.52, 1.90]	[1.72, 2.09]
AS arousal	72.85 (23.69)	80.62 (16.06)	67.36 (30.95)
	[66.08, 79.63]	[76.03, 85.21]	[58.51, 76.21]
AS valence	56.06 (24.98)	67.81 (24.40)	65.38 (32.48)
	[48.92, 63.20]	[60.83, 74.79]	[56.10, 74.67]
PANAS positive	24.32 (7.86)	26.30 (9.04)	26.93 (7.27)
	[22.07, 26.57]	[23.71, 28.88]	[24.73, 29.13]
PANAS negative	15.72 (6.12)	14.79 (5.73)	18.80 (7.16)
	[13.97, 17.47]	[13.15, 16.43]	[16.71, 20.89]
Baseline heart rate	80.50 (12.46)	77.93 (12.91)	—
	[76.86, 84.14]	[74.23, 81.62]	—
Experimental heart rate	80.88 (12.74)	81.49 (12.87)	—
	[77.12, 84.64]	[77.81, 85.17]	—

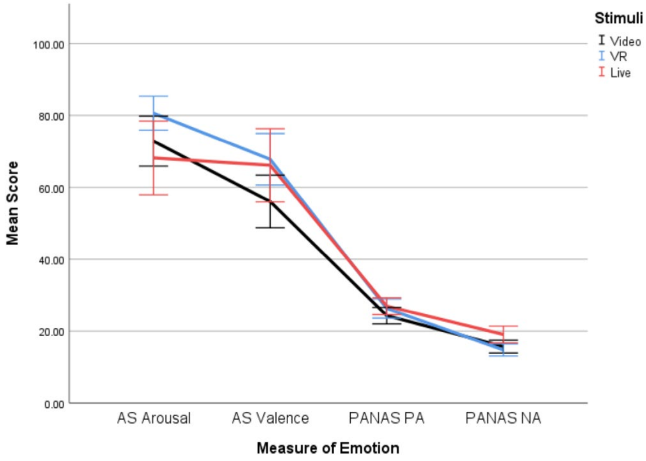

FIGURE 2 | Mean PANAS and AS emotion scores across conditions. Error bars represent 95% confidence intervals.

correlations. There were no significant correlations between HR and presence ($p > 0.05$, $r_s < 0.11$).

To address hypothesis 8, a mixed ANOVA was conducted to compare the HR (baseline and experimental) between the two groups. There was a significant main effect of time [$F(1, 89) = 11.61$, $p < 0.001$, $\eta_p^2 = 0.12$] and a significant interaction effect between time and group [$F(1, 89) = 9.07$, $p = 0.003$, $\eta_p^2 = 0.09$]. Post hoc analysis suggested that there was no significant difference in HR between the groups at either time point ($p > 0.05$). The video group experienced no significant change in HR between the time points ($p = 0.78$), the VR group did experience a significantly higher HR in the experimental phase compared to the baseline ($p < 0.001$), though this was a relatively small effect ($d = 0.28$).

3.2 | Exploratory Analysis

Following the initial, pre-registered data analysis, exploratory analysis was conducted to determine if the VR stimulus was more likely to induce VIMS than the video stimulus. A mixed ANOVA was conducted, comparing the SSQ scores between the groups before and after viewing. The VR group had a higher SSQ score before ($M = 76.52$, SD = 88.12) and after ($M = 87.54$, SD = 88.49) stimulus exposure than the video group (before: $M = 63.48$, SD = 74.33; after: $M = 51.81$, SD = 62.86). However, there was no significant main effect of time [$F(1, 92) = 0.002$, $p = 0.97$, $\eta p^2 < 0.001$] or a significant interaction effect between time and group [$F(1, 92) = 1.87$, $p = 0.18$, $\eta p^2 = 0.02$]. This suggests that the VR stimulus was no more likely to induce VIMS symptoms than the video. This supports the previous result that showed no difference in ITC-SOPI negative effect scores between the video and VR groups, which, like the SSQ, also includes physical symptoms.

The pre-registered analysis showed that the live group reported fewer correct and incorrect items than the other groups, potentially because they reported fewer items overall. A one-way ANOVA confirmed this, demonstrating a statistically significant difference in the total number of items reported in each group, $F(2, 138) = 15.41$, $p < 0.001$, $\eta^2 = 0.18$. Post hoc *t*-tests, employing the Bonferroni adjustment, showed that the live-event group recalled significantly fewer total items (video, $p < 0.001$; VR, $p < 0.001$) than the other groups, while the video and VR groups reported similar amounts of information ($p > 0.05$).

4 | Discussion

This study aimed to compare the experience of participant eyewitnesses who view a to-be-remembered event on video, in VR, or in a live event. In doing so, it sought to determine if VR offers a realistic eyewitness experience and if it could be a useful tool for researchers that provides the enhanced ecological validity of a live event while maintaining experimental control.

Participants who viewed the event in VR reported a similar level of spatial presence and engagement to the live-event group, suggesting that in this study, the reported sense of physical location and psychological involvement in the scene provided by VR is comparable to real life. This is not the case with video stimuli, which elicited significantly lower scores for both factors. Perhaps unsurprisingly, the live-event group had the highest ecological validity rating, with the VR experience rated as more real and lifelike than the video group. These findings suggest that VR offers an eyewitness experience that is more ecologically valid than video and may be psychologically similar to witnessing an event in person. This is in line with the review by Hepperle and Wölfel (2023), who found that VR offers experiences that are more similar to real life than video. Participants in the VR group experienced a significantly greater increase in physiological arousal than the video group, suggesting that the more realistic experience led to a more realistic physiological response. However, the presence measures did not correlate with HR, in contrast to a similar study conducted by Green et al. (2025). This is contrary to the current body of evidence, which suggests that HR is usually a reliable correlate of presence (Grassini and Laumann 2020). It may be the case that the stimulus in the current study was limited in its ability to evoke a physiological response, as experimental eyewitness research is restricted in how much stress it can intentionally cause participants (Chae 2010), and a non-crime event was used. Recording the HR of those in the live-event group, who may have believed the event to be real, might also shed some light on this issue.

Emotional experience was similar between the groups, suggesting that VR provides an emotional experience similar to real life. This reflects research by Chirico and Gaggioli (2019), who found that 360° videos presented in VR can elicit similar emotional responses to their real-world equivalents. The live-event group did, however, have a significantly greater negative affect compared to the other two groups. Participants were more likely to report feeling uncomfortable or shocked during the scene if it was viewed in person. It may be that though VR offers a realistic experience, it is enjoyable by nature (Glomb et al. 2024) which may reduce negative emotion. However, it is likely that as the live experience is the most ecologically valid, the negative emotions measured by the PANAS (such as being upset, irritable or hostile) were heightened in the inherently negative scene. The VR group reported a similar level of negative physiological effects to the live-event group; however, they were no more likely to experience VIMS than the video group. This suggests that using VR can elicit similar negative effects to real life, but when used in this way is unlikely to cause motion sickness.

There was no difference in free or cued recall accuracy between the groups, which supports some of the previous literature. (Pozzulo et al. 2008) found that there was no difference in recall between live and video stimuli, while Green et al. (2025) and Glomb et al. (2024) found that there was no difference between video and VR stimuli. This suggests that using VR does not positively or negatively impact recall and produces similar results to live events and should be considered as a methodologically valid alternative to traditional methods such as videos or live events. The live-event group, however, did report significantly fewer items in the free recall task (though were no less accurate). This is possibly due to the method of data entry, as they completed questionnaires on paper rather than on computer and handwrote responses, or could be due to perceived time pressure as the data collection took place during a lecture. Those in the video and VR groups, who knew the event was not real from the start, may have actively attempted to recall more information or felt under pressure to provide a higher level of detail due to demand characteristics.

Participants who view video stimuli have their attention directed by the researcher to focus on the event (Ihlebæk et al. 2003) and spend more time looking at people in the video (Foulsham et al. 2011) than they might in real life, which may impact encoding. Though participants could turn their heads to allocate their attention in VR, the view was uninterrupted and consistently clear, whereas the live-event group had varying perspectives depending on their location in the room, which may also help to explain the decrease in information reported. An increased sense of presence may imply increased levels of attention (Witmer and Singer 1998), which can impact recall. However, the implication is that attention is allocated to the virtual world rather than the real world, but may still be divided within the virtual environment, as it might be in the real world.

These findings suggest that VR may offer an experience that provides a similar sense of “being there” to a live eyewitness event, maintains experimental control, and is more ecologically valid than video stimuli. Further research is needed to confirm these findings; however, the authors suggest that eyewitness memory researchers consider using VR to display stimuli in their studies due to the methodological benefits outlined in this paper.

4.1 | Limitations and Future Direction

This study aimed to examine the differences in eyewitness memory research when using different stimulus mediums. Inferences have been made about witnessing a crime, though the authors acknowledge that the stimulus video did not include a crime scenario. However, it was an emotionally charged, forensically relevant event, which was designed to have many features of a crime (e.g., a clear perpetrator and victim). This is not an uncommon approach; for example, Flin et al. (1992) conducted a similar live event, which had actors knock over some items and engage in a heated argument in front of a lecture hall. Though this was a non-crime event, the findings were still directly relevant and applicable to eyewitness memory of crimes. There was also an expectation that participants in the live-event group would believe the staged event to be real, and the use of a non-crime event was intended to minimize the risk of participant distress or intervention. Unfortunately, this may have limited the ability of the stimulus to elicit levels of physical or emotional arousal similar to real life, which may explain the non-significant correlation between HR and presence, and the limited effect of stimulus type on HR. While VR may lead to a similar experience to live events, it is unlikely that viewers of a VR event would believe the events to be real. Therefore, future eyewitness memory research using VR should consider presenting crime events where possible.

All participants in the live-event group were final year psychology students, while the VR and video groups consisted mostly of psychology students, or other university students and staff. This narrow scope of participant recruitment is common in psychological research, though it can be difficult to generalize findings from students to the general population (Hanel and Vione 2016). This is also the case for eyewitness research; for example, Flowe et al. (2018) found that 69% of eyewitness identification studies used student populations. As the current study aims to evaluate existing methods against a new one, it is logical to recruit the most commonly used population; however, it should be noted that the findings may not apply to the general population. Beyond age and educational differences, university students may be more likely to have previous exposure to technologies like virtual reality than the general public. It should also be noted that the inter-rater reliability was mostly moderate, or poor. As mentioned in the results section, this could have been due to the subjective nature of the coding process or a need for enhanced training and guidance for the coders. Review of the second coder's data suggests that, despite the variation in coding, the same overall conclusions would have been drawn if it had been relied upon in the analysis. It should be noted that this issue is only related to the free recall findings, about which few strong claims are made in this paper. The authors acknowledge that the free recall finding at the edge of statistical significance, that the live group reported fewer incorrect items ($p = 0.03$), should be interpreted with caution; however, it is consistent with the overall amount of information provided by the group. The lack of difference in free recall between the video and VR group is consistent with previous research in the area (Glomb et al. 2024; Green et al. 2025), supporting the validity of the finding.

Another possible limitation is that the live-event group consisted solely of final year psychology students, while the other groups included participants from a wider pool. This likely means an unequal exposure to psychological research methods, as well as VR, which is integrated into their course. Due to this prior knowledge, they may have suspected the event to be staged, which could have impacted measures such as spatial presence, ecological validity, and emotional engagement. However, the high ecological validity rating on the ITC-SOPI, suggesting realism of experience, contradicts this. Where the live-event group may have believed it to be a real event, the other groups knew this not to be the case as they were watching a recording. However, the VR group reported having a very similar experience to the live-event group, so it seems that this was of little consequence to the findings.

The research would have benefited from recording the HR of the live-event group, for a real-world comparison to the other groups, to better gauge ecological validity. However, to ensure that the event was a surprise, and because of the number of simultaneous participants, this was not possible. The live-event participants witnessed the scene, and completed the questionnaires, as a largely unsupervised group and could have colluded. It could be the case that participants in this group with poor recall could have asked or copied their peers, inflating the performance of the group overall. The video and VR groups had a uniform experience in view angle and distance, close to the events, while the live-event group had a highly varied experience, with some participants sat near the back of the lecture room. Such variations in viewing distance can impact eyewitness memory (Semmler et al. 2018). Future research in this area could use live events presented to individuals or small groups, which would allow researchers to record HR, more easily monitor participants to reduce collusion and standardise viewing distance.

As it has been shown to offer an enhanced level of ecological validity, researchers could consider presenting to-be-remembered events in VR for a range of forensic purposes. For example, it would be appropriate for studies of eyewitness identification or weapon focus effect, though particular elements of the technology warrant further evaluation such as screen resolution. VR screen resolution can be substantially lower than videos, due to the stretching of the image over a 360° canvas, and this may impact the level of face or weapon detail available to participant-witnesses. VR provides researchers a unique opportunity to place participants in situations that would normally be difficult or dangerous to orchestrate in real life, while providing a level of realism superior to a video. For example, Maass and Köhnken (1989) stated that one of the key criteria that weapon focus research should meet is the presence of a weapon that participants would be fearful of. However, they also acknowledge that ethical considerations also limit the type of weapons that should be included, suggesting that knives or guns would not be appropriate. This restricts the ecological validity of the research, as they are commonly used in crimes. VR could provide an experience that offers enhanced ecological validity, compared to video, while still using knives or guns, as participants are not likely to believe them to be real. It would also be interesting to explore the differences in eye tracking between VR, live and video eyewitness scenarios. Ihlebæk et al. (2003) suggested that participants who watch video stimuli have their attention directed by the camera, and may pay attention to different aspects of the scene compared to participants who believe the event to be real. VR eye tracking, which is provided in a number of devices, would also allow eyewitness researchers to explore this as well as other factors such as weapon focus, with the benefits of VR's ecological validity.

The findings of this study suggest that VR may offer an ecologically valid and reliable eyewitness experience, which is highly similar to viewing an event in real life and is more realistic than traditional video stimuli. VR offers a method of displaying events that elicits real world responses, a sense of "being there" and recall accuracy representative of real life. Though this is the first study of its kind, future researchers could benefit from using VR to present to-be-remembered events in eyewitness memory research.

Author Contributions

Andrew D. Green: conceptualization, methodology, formal analysis, investigation, project administration, writing – original draft. **Andrew Clark:** conceptualization, methodology, supervision, writing – review and editing. **Joanne Rechdan:** writing – review and editing. **Andy Guppy:** writing – review and editing, supervision.

Acknowledgements

The authors would like to thank the research assistants who contributed to the coding of the stimulus video and participant responses, and the actors in the live event.

Ethics Statement

Ethical approval was obtained from the Research Centre for Applied Psychology at the University of Bedfordshire. The procedures used in this study adhere to the tenets of the Declaration of Helsinki, and all participants gave their informed consent prior to data collection.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are openly available in Open Science Framework at https://osf.io/zq7e6.

References

Bates, J. L., L. A. Ricciardelli, and V. A. Clarke. 1999. "The Effects of Participation and Presentation Media on the Eyewitness Memory of Children." *Australian Journal of Psychology* 51, no. 2: 71–76. https://doi.org/10.1080/00049539908255338.

Betella, A., and P. F. Verschure. 2016. "The Affective Slider: A Digital Self-Assessment Scale for the Measurement of Human Emotions." *PLoS One* 11, no. 2: e0148037. https://doi.org/10.1371/journal.pone.0148037.

Bradley, M. M., and P. J. Lang. 1994. "Measuring Emotion: The Self-Assessment Manikin and the Semantic Differential." *Journal of Behavior Therapy and Experimental Psychiatry* 25, no. 1: 49–59. https://doi.org/10.1016/0005-7916(94)90063-9.

Buckhout, R., D. Figueroa, and E. Hoff. 1975. "Eyewitness Identification: Effects of Suggestion and Bias in Identification From Photographs." *Bulletin of the Psychonomic Society* 6, no. 1: 71–74. https://doi.org/10.3758/BF03333151.

Busch, M., M. Lorenz, M. Tscheligi, C. Hochleitner, and T. Schulz. 2014. "Being There for Real: Presence in Real and Virtual Environments and Its Relation to Usability." In *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational*, 117–126. Association for Computing Machinery. https://doi.org/10.1145/2639189.2639224.

Cadet, L. B., and H. Chainay. 2020. "Memory of Virtual Experiences: Role of Immersion, Emotion and Sense of Presence." *International Journal of Human-Computer Studies* 144: 102506. https://doi.org/10.1016/j.ijhcs.2020.102506.

Ceccato, I., E. Ricci, C. Mazza, et al. 2024. "Influence of Stimuli Emotional Features and Typicality on Memory Performance: Insights From a Virtual Reality Context." *Psychological Research* 88, no. 1: 257–270. https://doi.org/10.1007/s00426-023-01850-8.

Chae, Y. 2010. "Application of Laboratory Research on Eyewitness Testimony." *Journal of Forensic Psychology Practice* 10, no. 3: 252–261. https://doi.org/10.1080/15228930903550608.

Chirico, A., and A. Gaggioli. 2019. "When Virtual Feels Real: Comparing Emotional Responses and Presence in Virtual and Natural Environments." *Cyberpsychology, Behavior, and Social Networking* 22, no. 3: 220–226. https://doi.org/10.1089/cyber.2018.0393.

Chun, M. M., and N. B. Turk-Browne. 2007. "Interactions Between Attention and Memory." *Current Opinion in Neurobiology* 17, no. 2: 177–184. https://doi.org/10.1016/j.conb.2007.03.005.

Deffenbacher, K. A., B. H. Bornstein, S. D. Penrod, and E. K. McGorty. 2004. "A Meta-Analytic Review of the Effects of High Stress on Eyewitness Memory." *Law and Human Behavior* 28: 687–706. https://doi.org/10.1007/s10979-004-0565-x.

Dibbets, P., and M. A. Schulte-Ostermann. 2015. "Virtual Reality, Real Emotions: A Novel Analogue for the Assessment of Risk Factors of Post-Traumatic Stress Disorder." *Frontiers in Psychology* 6: 681. https://doi.org/10.3389/fpsyg.2015.00681.

Felnhofer, A., O. D. Kothgassner, M. Schmidt, et al. 2015. "Is Virtual Reality Emotionally Arousing? Investigating Five Emotion Inducing Virtual Park Scenarios." *International Journal of Human-Computer Studies* 82: 48–56. https://doi.org/10.1016/j.ijhcs.2015.05.004.

Flin, R., J. Boon, A. Knox, and R. Bull. 1992. "The Effect of a Five-Month Delay on Children's and Adults' Eyewitness Memory." *British Journal of Psychology* 83, no. 3: 323–336. https://doi.org/10.1111/j.2044-8295.1992.tb02444.x.

Flowe, H. D., A. Carline, and N. Karoğlu. 2018. "Testing the Reflection Assumption: A Comparison of Eyewitness Ecology in the Laboratory and Criminal Cases." *International Journal of Evidence and Proof* 22, no. 3: 239–261. https://doi.org/10.1177/1365712718782996.

Foster, J. L., T. Huthwaite, J. A. Yesberg, M. Garry, and E. F. Loftus. 2012. "Repetition, Not Number of Sources, Increases Both Susceptibility to Misinformation and Confidence in the Accuracy of Eyewitnesses." *Acta Psychologica* 139, no. 2: 320–326. https://doi.org/10.1016/j.actpsy.2011.12.004.

Foulsham, T., E. Walker, and A. Kingstone. 2011. "The Where, What and When of Gaze Allocation in the Lab and the Natural Environment." *Vision Research* 51, no. 17: 1920–1931. https://doi.org/10.1016/j.visres.2011.07.002.

Glomb, K., P. Piotrowski, and I. A. Romanowska. 2024. "It Is Not Real Until It Feels Real: Testing a New Method for Simulation of Eyewitness Experience With Virtual Reality Technology and Equipment." *Behavior Research Methods* 56, no. 5: 4336–4350. https://doi.org/10.3758/s13428-023-02186-2.

Grassini, S., and K. Laumann. 2020. "Questionnaire Measures and Physiological Correlates of Presence: A Systematic Review." *Frontiers in Psychology* 11: 349. https://doi.org/10.3389/fpsyg.2020.00349.

Green, A. D., A. Clark, M. Pitchford, and A. Guppy. 2025. "Viewing Mock Crimes in Virtual Reality Increases Presence Without Impacting Memory." *Behavior Research Methods* 57, no. 3: 1–15. https://doi.org/10.3758/s13428-024-02575-1.

Hanel, P. H., and K. C. Vione. 2016. "Do Student Samples Provide an Accurate Estimate of the General Public?" *PLoS One* 11, no. 12: e0168354. https://doi.org/10.1371/journal.pone.0168354.

Hepperle, D., and M. Wölfel. 2023. "Similarities and Differences Between Immersive Virtual Reality, Real World, and Computer Screens: A Systematic Scoping Review in Human Behavior Studies." *Multimodal Technologies and Interaction* 7, no. 6: 56. https://doi.org/10.3390/mti7060056.

Ihlebæk, C., T. Løve, D. Erik Eilertsen, and S. Magnussen. 2003. "Memory for a Staged Criminal Event Witnessed Live and on Video." *Memory* 11, no. 3: 319–327. https://doi.org/10.1080/09658210244000018.

Innocence Project. 2025. "Explore the Numbers: Innocence Project's Impact." https://innocenceproject.org/exonerations-data/.

Kennedy, R. S., N. E. Lane, K. S. Berbaum, and M. G. Lilienthal. 1993. "Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness." *International Journal of Aviation Psychology* 3, no. 3: 203–220. https://doi.org/10.1207/s15327108ijap0303_3.

Kensinger, E. A., and S. Corkin. 2004. "Two Routes to Emotional Memory: Distinct Neural Processes for Valence and Arousal." *Proceedings of the National Academy of Sciences* 101, no. 9: 3310–3315. https://doi.org/10.1073/pnas.0306408101.

Kensinger, E. A., and D. L. Schacter. 2006. "Processing Emotional Pictures and Words: Effects of Valence and Arousal." *Cognitive, Affective, and Behavioral Neuroscience* 6, no. 2: 110–126. https://doi.org/10.3758/CABN.6.2.110.

Kim, K., K. K. Park, and J. H. Lee. 2014. "The Influence of Arousal and Expectation on Eyewitness Memory in a Virtual Environment." *Cyberpsychology, Behavior, and Social Networking* 17, no. 11: 709–713. https://doi.org/10.1089/cyber.2013.0638.

Kisker, J., T. Gruber, and B. Schöne. 2021. "Behavioral Realism and Lifelike Psychophysiological Responses in Virtual Reality by the Example of a Height Exposure." *Psychological Research* 85, no. 1: 68–81. https://doi.org/10.1007/s00426-019-01244-9.

Kloft, L., H. Otgaar, A. Blokland, et al. 2020. "Cannabis Increases Susceptibility to False Memory." *Proceedings of the National Academy of Sciences* 117, no. 9: 4585–4589. https://doi.org/10.1073/pnas.1920162117.

Kober, S. E., and C. Neuper. 2012. "Using Auditory Event-Related EEG Potentials to Assess Presence in Virtual Reality." *International Journal of Human-Computer Studies* 70, no. 9: 577–587. https://doi.org/10.1016/j.ijhcs.2012.03.004.

Kothgassner, O. D., and A. Felnhofer. 2020. "Does Virtual Reality Help to Cut the Gordian Knot Between Ecological Validity and Experimental Control?" *Annals of the International Communication Association* 44: 1–9. https://doi.org/10.1080/23808985.2020.1792790.

LaPaglia, J. A., M. M. Wilford, J. R. Rivard, J. C. Chan, and R. P. Fisher. 2014. "Misleading Suggestions Can Alter Later Memory Reports Even Following a Cognitive Interview." *Applied Cognitive Psychology* 28, no. 1: 1–9. https://doi.org/10.1002/acp.2950.

Lessiter, J., J. Freeman, E. Keogh, and J. D. Davidoff. 2001. "A Cross-Media Presence Questionnaire: The ITC Sense of Presence Inventory." *Presence Teleoperators and Virtual Environments* 10, no. 3: 282–297. https://doi.org/10.1162/105474601300343612.

Loftus, E. F. 2005. "Planting Misinformation in the Human Mind: A 30-Year Investigation of the Malleability of Memory." *Learning and Memory* 12, no. 4: 361–366. https://doi.org/10.1101/lm.94705.

Loftus, E. F., D. G. Miller, and H. J. Burns. 1978. "Semantic Integration of Verbal Information Into a Visual Memory." *Journal of Experimental Psychology: Human Learning and Memory* 4, no. 1: 19–31. https://doi.org/10.1037/0278-7393.4.1.19.

Lombard, M., and T. Ditton. 1997. "At the Heart of It All: The Concept of Presence." *Journal of Computer-Mediated Communication* 3, no. 2: JCMC321. https://doi.org/10.1111/j.1083-6101.1997.tb00072.x.

Maass, A., and G. Köhnken. 1989. "Eyewitness Identification: Simulating the "Weapon Effect"." *Law and Human Behavior* 13, no. 4: 397–408. https://doi.org/10.1007/BF01056411.

Metcalfe, J., J. C. Brezler, J. McNamara, G. Maletta, and M. Vuorre. 2019. "Memory, Stress, and The Hippocampal Hypothesis: Firefighters' Recollections of the Fireground." *Hippocampus* 29, no. 12: 1141–1149. https://doi.org/10.1002/hipo.23128.

Nisenfeld, S. 2003. "Using Reality to Evaluate the ITC Presence Questionnaire." https://cs.brown.edu/research/pubs/theses/masters/2003/nisenfeld.pdf.

Nyman, T. J., J. Antfolk, J. M. Lampinen, J. Korkman, and P. Santtila. 2020. "Eyewitness Identifications After Witnessing Threatening and Non-Threatening Scenes in 360-Degree Virtual Reality (Or 2D) From First and Third Person Perspectives." *PLoS One* 15, no. 9: e0238292. https://doi.org/10.1371/journal.pone.0238292.

Pozzulo, J. D., C. Crescini, and T. Panton. 2008. "Does Methodology Matter in Eyewitness Identification Research?: The Effect of Live Versus Video Exposure on Eyewitness Identification Accuracy." *International Journal of Law and Psychiatry* 31, no. 5: 430–437. https://doi.org/10.1016/j.ijlp.2008.08.006.

Reeves, B. 1991. "Being There: Television as Symbolic Versus Natural Experience." Institute for Communication Research, Stanford University, Stanford, CA.

Rubínová, E., R. J. Fitzgerald, S. Juncu, E. Ribbers, L. Hope, and J. D. Sauer. 2021. "Live Presentation for Eyewitness Identification Is Not Superior to Photo or Video Presentation." *Journal of Applied Research in Memory and Cognition* 10, no. 1: 167–176. https://doi.org/10.1016/j.jarmac.2020.08.009.

Schöne, B., M. Wessels, and T. Gruber. 2019. "Experiences in Virtual Reality: A Window to Autobiographical Memory." *Current Psychology* 38: 715–719. https://doi.org/10.1007/s12144-017-9648-y.

Schubert, T., F. Friedmann, and H. Regenbrecht. 2001. "The Experience of Presence: Factor Analytic Insights." *Presence Teleoperators and Virtual Environments* 10, no. 3: 266–281. https://doi.org/10.1162/105474601300343603.

Semmler, C., J. Dunn, L. Mickes, and J. T. Wixted. 2018. "The Role of Estimator Variables in Eyewitness Identification." *Journal of Experimental Psychology: Applied* 24, no. 3: 400–415. https://doi.org/10.1037/xap0000157.

Slater, M. 2018. "Immersion and the Illusion of Presence in Virtual Reality." *British Journal of Psychology* 109, no. 3: 431–433. https://doi.org/10.1111/bjop.12305.

Stanny, C. J., and T. C. Johnson. 2000. "Effects of Stress Induced by a Simulated Shooting on Recall by Police and Citizen Witnesses." *American Journal of Psychology* 113, no. 3: 359–386. https://doi.org/10.2307/1423364.

van Dammen, L., T. T. Finseth, B. H. McCurdy, et al. 2022. "Evoking Stress Reactivity in Virtual Reality: A Systematic Review and Meta-Analysis." *Neuroscience and Biobehavioral Reviews* 138: 104709. https://doi.org/10.1016/j.neubiorev.2022.104709.

Watson, D., L. A. Clark, and A. Tellegen. 1988. "Development and Validation of Brief Measures of Positive and Negative Affect: The PANAS Scales." *Journal of Personality and Social Psychology* 54, no. 6: 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063.

Wells, G. L., and S. D. Penrod. 2011. "Eyewitness Identification Research: Strengths and Weaknesses of Alternative Methods." In *Research Methods in Forensic Psychology*, edited by B. Rosenfeld and S. D. Penrod, 237–256. Wiley and Sons.

Wilford, M. M., J. C. Chan, and S. J. Tuhn. 2014. "Retrieval Enhances Eyewitness Suggestibility to Misinformation in Free and Cued Recall." *Journal of Experimental Psychology: Applied* 20, no. 1: 81–93. https://psycnet.apa.org/doi/10.1037/xap0000001.

Witmer, B. G., and M. J. Singer. 1998. "Measuring Presence in Virtual Environments: A Presence Questionnaire." *Presence* 7, no. 3: 225–240. https://doi.org/10.1162/105474698565686.

Wright, A. M., and R. E. Holliday. 2007. "Enhancing the Recall of Young, Young–Old and Old–Old Adults With Cognitive Interviews." *Applied Cognitive Psychology* 21, no. 1: 19–43. https://doi.org/10.1002/acp.1260.

Wulff, A. N., and A. K. Thomas. 2021. "The Dynamic and Fragile Nature of Eyewitness Memory Formation: Considering Stress and Attention." *Frontiers in Psychology* 12: 666724. https://doi.org/10.3389/fpsyg.2021.666724.