The Effect of Augmented Reality on Involuntary Autobiographical Memory

Leana Petiot
Inria, Univ. Bordeaux, CNRS
Bordeaux, France
leana.petiot@inria.fr

Hélène Sauzeon
Inria, Univ. Bordeaux, CNRS, BPH
Bordeaux, France
helene.sauzeon@inria.fr

Pierre Dragicevic
Inria, Univ. Bordeaux, CNRS
Bordeaux, France
pierre.dragicevic@inria.fr

Abstract

We know little about the impact of augmented reality (AR) on human cognition, particularly regarding involuntary autobiographical memory (IAM). IAMs are spontaneous recollections of personal events, ubiquitous in daily life but under-researched in both psychology and human-computer interaction. We first discuss the potential opportunities and risks of replacing conventional displays with AR to increase the likelihood of IAMs. We then report on a study investigating whether stimuli displayed on the same mobile device using video-see-through AR are more likely to resurface than those shown with a simple 3D viewer. We found that AR elicits approximately twice as many IAMs in controlled settings with immediate re-exposure to contextual cues, but no measurable effect was found in everyday settings with delayed re-exposure. Therefore, AR can enhance IAMs, but its effects may be modest and short-lived in most cases. Nevertheless, future studies could reveal stronger effects of AR in other settings. Supplementary material is available at osf.io/gscpv/.

CCS Concepts

• **Human-centered computing → Empirical studies in HCI**.

Keywords

Involuntary Autobiographical Memory, Augmented Reality

ACM Reference Format:
Leana Petiot, Hélène Sauzeon, and Pierre Dragicevic. 2025. The Effect of Augmented Reality on Involuntary Autobiographical Memory. In *CHI Conference on Human Factors in Computing Systems (CHI '25), April 26–May 01, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 20 pages. https://doi.org/10.1145/3706598.3713922

Figure 1: Illustration of a mobile app like IKEA Place [62], allowing customers to virtually place products (here, a bedside lamp) in their home before buying them. ***Image by the authors, CC-BY.***

1 Introduction

Several applications of augmented reality (AR) have become commonplace. For example, more and more web stores use handheld AR to let users preview how products would look like in their homes (see Figure 1). For such tasks where AR is clearly useful, AR may progressively replace traditional ways of displaying digital information. But could this increase risks of user manipulation? For example, imagine a company that requires people to watch product ads in AR, in exchange for some free service. A person might see an ad for a bedside lamp like the one in Figure 1, while playing a game on their bed. Each time this person lies on their bed, will they be more likely to spontaneously recall the lamp than if they saw it on a regular ad on their phone? In other words, can the memorability of AR imagery be instrumentalized to influence purchasing decisions, or manipulate thoughts and behavior in other ways?

In psychology, a spontaneous recollection as in the sofa example is called an involuntary autobiographical memory, or IAM [12].[1] It is a commonly-occurring form of memory characterized by an effortless recall triggered by cues, such as a thought or an object in the environment. This type of memory is not easily controllable by individuals, which makes it particularly vulnerable to manipulation. But despite their importance, there is little research on IAMs, which are a recently developed topic in psychology compared to other forms of memory. Meanwhile, there is almost no research on IAMs in the area of human-computer interaction (HCI). In particular, research on extended reality (XR)[2] has largely focused on voluntary memory (e.g., [49, 61, 74]). In this article, we study the effect of AR on involuntary memories. More specifically, we look at whether displaying digital content in AR elicits more involuntary memories of the content than using conventional computer displays. Not only are AR experiences generally more engaging and therefore potentially more memorable, AR also has the capacity to embed digital content directly into physical environments, which could increase the likelihood that the physical environment (e.g., the

[1]Like previous work [12], we use IAM both to refer to the memory system and to recollected content (i.e., memories).
[2]Like others [70], we use XR as an umbrella term to refer to augmented reality (AR), virtual reality (VR), and mixed reality (MR).

This work is licensed under a Creative Commons Attribution 4.0 International License.
CHI '25, Yokohama, Japan
© 2025 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-1394-1/25/04
https://doi.org/10.1145/3706598.3713922

bedroom in our example) acts as a cue for an involuntary memory (e.g., the sofa). Although computer devices (e.g., a smartphone) and digital content (e.g., a web search page) can themselves act as memory cues, regular computer users are constantly exposed to them and they are associated with a wide range of content, which potentially makes them weaker memory cues [45, 92].

If AR really does promote involuntary memories, the implications can be far-reaching, both in terms of potential risks (e.g., abusive advertisement, as we have already seen) and opportunities—we will go through both negative and positive scenarios in subsection 2.1. In any case, if AR becomes mainstream and if application designers and content providers start relying on AR instead of regular displays to create a stronger impression on users, it is essential that we are able to understand and quantify the impact of AR on involuntary memory. Doing so will allow us to get a much better sense of the power of such immersive displays compared to traditional displays. However, as far as we know, the question of whether AR can cause more involuntary memories than conventional displays has so far never been empirically tested.

This article reports two experiments. In the first experiment, participants were asked to walk along a path with a smartphone, and to examine a sequence of visual stimuli (i.e., 3D objects such as a dartboard or a sculpture) laid across that path. A stimulus was shown on the smartphone either using a regular 3D model viewer or using video-see-through AR, depending on the condition. Then, participants were asked to walk along the same path again without the stimuli, and report any wandering thought while performing a so-called "vigilance task" (in our case an auditory counting task), i.e., a monotonous task designed to encourage IAMs and minimize voluntary recall [16, 77]. During this task, participants were oblivious to the true purpose of the experiment. Participants were then asked to report any involuntary memories of the stimuli during a 4-day follow-up diary study outside the lab. We found evidence that content shown in AR led to more involuntary memories in the lab, but the results were inconclusive for the diary study. We conducted a second experiment similar to the first, except participants were told the true purpose of the experiment just before the audio task and were specifically asked to report involuntary memories of stimuli. This second experiment led to similar results.

To sum up, this article brings three major contributions: (1) scenarios illustrating how AR can be used to influence involuntary memory in real life, (2) the first experimental paradigm for investigating the effect of AR on involuntary memories which combines methodology from psychology and HCI, and comes with comprehensive replication material, and (3) first evidence for a facilitating effect of AR on involuntary memory.

2 Background

In this section, we begin by presenting real and fictional examples meant to illustrate how involuntary autobiographical memory (IAM) might be affected by AR experiences, and the possible consequences. We then summarize what we know about IAMs from a cognitive psychology perspective, including the methodologies used in the past to elicit and to quantify IAMs. Finally, we provide a brief overview of studies of memory in XR settings.

2.1 Motivating Scenarios

Here we go through hypothetical examples to illustrate the practical relevance of IAM in AR systems. Most examples are inspired by existing apps and can be implemented with current technologies, but we will also cover speculative examples that ignore current technological limitations in order to emphasize the potential impact AR may have on IAMs in the future.

Increasing customer sales. We already mentioned the IKEA Place app in the introduction (see Figure 1). More and more online stores offer a similar feature, including now Amazon [1]; Similar to IKEA Place, Amazon offers to preview some of its products in-place using AR, arguing that it reduces the risk of making a bad purchase. Customers can even try on virtual sneakers (see Figure 2-a). However, as we mentioned in the introduction, the use of AR could make it more likely that the previewed products spontaneously come back to the users' mind at a later time, for example when they discuss clothes with their friends, or even when they simply look at other people's shoes or at their own feet. If such recalls are recurring, they could help impress the product in people's minds and increase their desire for it. If AR does indeed have this effect and if companies discover that they can increase their profits with such a strategy, they might start using AR in order to influence purchasing decisions. People's freedom and autonomy could be impacted, especially if AR ads become ubiquitous, as in Keiichi Matsuda's dystopian short movies [54].

Raising public awareness. In contrast with the previous example, it is possible to imagine AR being used for broadly desirable purposes. For example, during the International Women's Day in 2023, Snap AR and the snapwomen association launched an initiative called *"March 8, 8 Women"*, where virtual models of important women were shown in French cities next to men statues [2]. Figure 2-b shows an example where a virtual statue of Simon Veil is displayed alongside the physical statue of Charles de Gaulle in Paris. The goal of the campaign was to draw attention to historical women, as people represented in French public spaces are only 10% women. Although it might not have been the designers' conscious intention, it is possible that after having seen one of the women in AR, users will sometimes spontaneously remember her when they are exposed to relevant cues (e.g., when passing through the same square or when thinking about de Gaulle). This in turn could help reinforce the importance of historical women in the user's mind. We can imagine public institutions or NGOs launching initiatives like these, that leverage AR to increase public awareness on key social issues. As another hypothetical example, it has been speculated that if an AR documentary visually embeds a refugee camp in somebody's backyard, the viewer may *"create a mental association and remember the refugees each time they see (or even think about) their backyard"* [27]. However, it is clear that even in the hands of public institutions, such strategies could serve both positive and negative political agendas. For example, in 2022 a viral video was posted simulating a Russian bombing of Paris [48]. Although the purpose of the video was presumably to increase public support for Ukraine, it has also highlighted risks of public manipulation. The simulated bombing was a simple video posted on social media, but it is possible to imagine an AR version that creates a strong impression on people, possibly even causing intrusive memories.

Figure 2: a) A user is trying virtual shoes on the Amazon web store [1] ; b) A virtual statue of French politician Simone Veil is added alongside the statue of Charles de Gaulle [2] ; c) A person uses handheld AR to visualize their water consumption, as if the water they used was flooding their bathroom [3]. ***Credits: a) Screenshot by the authors, CC-BY ; b) Image created by the authors for illustration purposes, CC-BY ; c) Image from [3], used with author permission.***

Promoting positive behavior. As another example of potentially positive application, AR could be used to promote healthy, pro-social, or pro-environmental behavior by the way of involuntary memories. Researchers have already started to explore the use of AR as an eco-feedback medium, for instance by showing virtual waste accumulations directly in users' physical surroundings [3]. For example, it is hard to get a visceral sense of our water consumption, as waste water rapidly disappears in water pipes once used. As illustrated in Figure 2-c, concrete AR visualizations can be used to convey water use in the user's surroundings as if waste water was flooding their physical environment. Again, besides the emotional response that such a concrete representation may elicit, having seen it embedded in a meaningful physical environment may facilitate its spontaneous recollection, especially when the user returns to the bathroom. This in turn could act as a reminder to consume less water by, e.g., taking a shorter shower. If such persuasive techniques are indeed effective, we could imagine them being employed to help people exercise more (e.g., by associating visual imagery related to exercising to relevant cues in the environment), or even help them act more positively towards other individuals ; For example, in the AR refugee camp example we mentioned previously, the increase in the frequency of spontaneous memories could increase the likelihood of charitable giving. Similar strategies could potentially even be used to help people curb unwanted behavior, by creating associations between dissuading imagery and thoughts, places or objects that could trigger the unwanted behavior, such as a pack of cigarettes, a bottle of wine, or snack items.

In all the hypothetical examples we have discussed, user consent is key to ethical use: AR imagery that can potentially "hack" somebody's mind should only be enabled if the user deliberately chooses to. However, fully informed consent requires that users understand the potential impact of AR on their cognition, especially on aspects on which they have little control such as their involuntary autobiographical memory (IAM). Currently, even scientists do not have such an understanding, so we need studies. Moreover, since it is impossible to guarantee that such techniques will always be deployed with people's consent, it is important we know how powerful AR can be at manipulating IAM, so we know how much energy and resources should be deployed in preventing misuse.

2.2 Involuntary Autobiographical Memory

We start by explaining what autobiographical memory (AM) is. Then, we explain and discuss involuntary autobiographical memory (IAM) more specifically.

2.2.1 Autobiographical memory. Simply speaking, autobiographical memories (AMs) are memories of our past experiences. Examples of *non*-autobiographical memories include semantic memory (e.g., remembering the capital of Italy) [72], procedural memory (e.g., remembering how to ride a bike) [88], and working memory (e.g remembering a phone number long enough to dial it) [5]. According to Martin A. Conway, a pioneering researcher on the topic, the term autobiographical memory specifically refers to *"our memory for specific episodes (episodic memory) and to our conceptual, generic, and schematic knowledge of our lives (autobiographical knowledge)"* [25]. According to him, AM is intricately linked to our "self", i.e., to our goals and self-images [26]; AM modulates the way in which we see our past selves, we interpret our present selves and we project our future selves, while our selves (including our goals and self-images) regulate the accessibility of our AMs [26]. Therefore, if AR can impact AM, it has the potential of impacting the very basis of our identity.

2.2.2 Voluntary vs. involuntary recall. In memory research, it is common to distinguish between *voluntary autobiographical memory* (VAM) and *involuntary autobiographical memory* (IAM) [9]. VAM refers to the cognitive processes involved when intentionally trying to recall past experiences—for example, when I try to remember the last time I ate at an Italian restaurant. In contrast, IAM refers to the processes involved when past experiences surface spontaneously [9]—for example, someone brings up Italian food during a conversation, and I immediately remember the last time I had dinner in an Italian restaurant. VAM and IAM therefore differ by the presence or absence of a conscious intention and effort to recall past experiences [7]. Some have questioned this sharp distinction, arguing that there is a continuum between voluntary and involuntary recall [60]. Nevertheless, the VAM/IAM distinction is a convenient simplification and we will use it, consistently with the literature. Importantly, both VAM and IAM involve conscious experiences, in contrast with *implicit memory* where recall is both involuntary and unconscious [34], and which is outside our scope–for example, when I see spaghetti and I instinctively know how to twirl my fork to eat it. Notably, studies have shown that IAMs are more prevalent in daily life than VAMs [68, 69], despite the lower research attention they have received.

2.2.3 Cues. IAMs are triggered by *cues* [9], which can be either external or internal. *External cues* are elements or aspects of our environment that are perceived by our senses, while *internal cues* are inner experiences such as thoughts or emotions [9]. Section 2.1 gave some examples: a person can remember Simone Veil when passing through the same square (external cue) or when thinking about de Gaulle (internal cue). *Ecphoric power* refers to the capacity of a cue to evoke a memory [87]: the higher a cue's ecphoric power, the higher the probability that it will trigger the IAM. Multiple studies have looked at determinants of ecphoric power. We know for example that the more elements a cue shares with the information stored in memory, the higher its ecphoric power [11, 87]. For example, thinking about De Gaulle may bring up the memory of encountering a virtual model of Simone Veil because a statue of him was present when the event was encoded; While thinking about another military leader might bring up the same memory, it is less likely to happen. Furthermore, verbal cues (e.g., the phrase "a glass of wine") were found to have higher ecphoric power than pictorial cues (e.g., a picture of a glass of wine) [55], and abstract cues (related to thought and language) were found to be more effective than sensory cues (e.g., sound, smell, sight) and internal states (e.g., feeling hungry or happy) [52]. Finally, according to the *fan effect* [16] (also called *overload theory* [92]), a cue has less ecphoric power if it is familiar and had been already associated with lots of personal memories. For example, passing by the statue of De Gaulle is much less likely to trigger the memory of Simone Veil if the viewer is a Parisian who walks there every day than if the viewer is a traveler who returns there for the second time.

2.2.4 Other factors affecting recall. IAMs (and AMs more generally) need to pass a *awareness threshold* to surface to consciousness, and this threshold is lower for memories that are sensory, emotionally intense, personally relevant, or unusual [60]. It also varies across individuals. The threshold can be lowered in experimental settings, for example when a participant is asked to retrieve a specific type of memory [60]. Finally, the more time passes after an event has been encoded in memory, the less likely it is to spontaneously resurface [29, 39, 85].

2.3 Measuring IAMs

IAMs are difficult to measure because they cannot be reliably elicited and cannot be objectively observed. However, several studies have attempted to measure IAMs. This section summarizes the methods that have been used.

2.3.1 Out-of-lab methods. A common technique for measuring IAMs in real-life settings is the *diary method* or *journaling technique*, which consists of asking participants to record their IAMs as they occur during the day. Despite the limited reliability of this method (e.g., participants may forget to record IAMs), it can help estimate the frequency of IAMs in ecological settings, and has proven helpful to study the phenomenological characteristics of IAMs [13, 50]. With modern technology, the diary method had increasingly relied on mobile phones. A study comparing mobile phones with paper diaries found that phones are less likely to be forgotten at home and are less prone to recording delays [46]. However, people report less IAMs on mobile phones, possibly because of a fan effect (see subsubsection 2.2.3): since mobile phones are highly familiar objects that serve many purposes, they may be a less powerful reminder of the experimental task than paper diaries, unless their appearance is altered [45]. Some diary studies use notifications to maintain participant engagement. For example, one study asking participants to report everyday memory failures (e.g., forgetting something) used daily SMS reminders to keep them engaged in the study [46]. Similarly, a study focusing on Involuntary Musical Imagery (INMI, or earworms—tunes that arise spontaneously and occupy one's thoughts) [31] employed the experience sampling method (ESM), sending text notifications at random times to prompt participants to report any occurrences of INMI. Another study investigating AMs and prospective memories (i.e., remembering to do something) used ESM, making random automated phone calls to prompt participants to report immediate or recent memories [32]. However, some researchers argue that ESM is inappropriate for studying IAMs [45]. Since people tend to quickly forget they had an IAM, they may forget to report many IAMs or falsely report VAMs as IAMs. Finally, a complement to the diary method is the use of *retrospective questionnaires and surveys* [14, 15], which ask participants to report the frequency of IAMs they had in the past. However, these methods are limited, again because people's memory of an IAM quickly fades out if it is not immediately recorded [9, 14].

2.3.2 Lab methods. Several methods have been used to elicit and measure IAMs in the laboratory. The *word-cue method* consists of first showing participants a series of phrases (e.g., being at the movies, sitting in class), and then asking them if they had any recollection of past experiences [51]. A variant is the *vigilance task paradigm*, which consists of pairing a word-cue task with a monotonous task [77]. For example, a computer screen displays a sequence of lines and occasionally word-cues, and the participant is asked to both detect vertical lines and report any IAM on-the-fly. The vigilance task is meant to elicit a state of mind-wandering known to facilitate IAMs [9], and to minimize voluntary recalls caused by retrospection. The vigilance task paradigm has been extensively used in IAM research [24, 44, 77]. Because participants are explicitly asked to monitor IAMs, their awareness threshold is reduced which may increase IAMs compared to more ecological settings [90]. To gain more ecological validity, some studies have asked participants to report any unrelated thought during the vigilance task, without disclosing that the goal was to measure IAMs [55].

2.4 Memory in XR Environments

Previous research on memory and XR can be roughly divided into *i)* research that uses XR *as a tool to study memory* in settings that are more ecological than conventional computer displays and more controlled than real life, and *ii)* research that uses XR *as an object of study*, e.g., by looking at the effect of XR technology on memory, or at ways in which XR technology can be used to enhance memory. We provide a brief summary of both lines of research.

2.4.1 XR as a Tool to Study Memory. Psychology research has shown interest in XR technologies, as they can mimic aspects of real life (thus providing some ecological validity) while providing the experimental control necessary for lab experiments. In particular, VR has been used to study autobiographical memory (AM) [66, 71, 76], see [83] for a review. Some studies have looked at how people encode AMs in VR compared to conventional computer displays. According to Schöne [78], VR experiences are encoded in a similar way to real-life personal experiences and can evoke rich AMs, whereas 2D videos are encoded as isolated episodic events [40]. However, most studies on AMs measure voluntary autobiographical memories (VAMs). We only know of two studies measuring IAMs, where VR was used to emulate the conditions for post-traumatic stress disorder (PTSD) [53, 63]. Contrary to VR, AR has not been a popular tool for studying memory, although some researchers have discussed possible advantages of AR—e.g., people can be studied while being in the real word, and stimuli can be easily superimposed on the physical environment [4].

2.4.2 Effects of XR on Memory. Some research has looked at how to use XR to *enhance human memory*. In human-computer interaction, AR concepts and prototypes have been proposed that leverage the "memory palace" mnemonic technique [74, 95]. Educational sciences have similarly looked at whether AR can be used to improve learning [22, 38, 97]. In particular, studies have found that AR can improve short-term knowledge acquisition and retention compared to traditional learning materials [84], and can improve memory for the location of objects compared to photographs [61]. Another stream of work has looked at whether XR can be used to *manipulate human memory*. Papers focusing on XR ethics have discussed possible scenarios of memory manipulation, including the use of XR to create distorted or false memories [19, 80]. They warn that the rapid development of XR techniques could increase this concern and produce more false memories that could potentially influence behavior. Other papers have experimentally studied false memories in VR by looking at source confusion, i.e., how often people tend to remember virtual objects as real objects and vice versa [18, 37, 75]. For example, in a 2001 study [37], 18% of the stimuli were attributed to the wrong source, and a 2024 replication [18] found 23% of misattributed stimuli. There is much less work on the topic of false memories in AR than in VR, with the exception of a study by Fernandes et al. [30], finding that source confusion exists in AR but decreases when people are invited to manipulate the (virtual and physical) stimuli. This result aligns with previous research findings that actively controlling changes in view can enhance memory [79]. The study by Fernandes comes close to our work by looking at the effect of AR on AM, but the focus is again on VAM, where recall is voluntary. To our knowledge, our study is the first to investigate the effect of AR on IAM.

2.4.3 Digital Mementos. Digital mementos are a stream of research at the intersection of IAM research and human-computer interaction (HCI). Mementos are objects such as photos that people use as cues to regularly remind themselves of personally-relevant events, persons, or places [65], or that act as memory aids [20]. A qualitative study found that more memories arise from physical mementos (e.g., personal objects) than digital mementos (e.g., social media posts, e-mails, or digital photos) [89]. In addition, personal objects that people repeatedly use or see during their daily activities (e.g., inherited cutlery, a guitar) can be powerful IAM cues, but digital mementos generally do not possess such a quality [89]. Meanwhile, a field study found that people tend to remember and value repeated experiences and routines (e.g., a mother's lullaby), something that mementos like photos do not support well since they tend to focus on single events [59]. While these findings provide valuable insights into the design of digital mementos for cueing IAMs, this research does not cover the effect of XR on IAM.

3 Experiment 1

This first experiment consisted of two parts: a *lab part* and a *diary part* (see Figure 3). First, during a *visual task* phase, participants were invited to the lab and asked to walk along a path with a smartphone to examine visual stimuli shown in AR or non-AR (see Figure 4). Then, during an *audio task* phase, they repeated the path while performing an auditory vigilance task, and were asked to report any wandering thought. At this stage, participants were unaware of the experiment's true purpose. Then, during a 4-day follow-up diary study outside the lab (the *diary task* phase), they were asked to record any involuntary memories of the experimental stimuli. The experiment was approved by the Inria ethical review board (COERLE, approval n°13953) and pre-registered on OSF[3].

3.1 Procedure

The participant was welcomed to our lab, where they were given a consent form and general instructions about the experiment. In order to increase the validity of our IAM measurements and minimize voluntary recall, the true purpose of the study was initially concealed from to the participant. The stated purpose was to study new ways of enhancing visitors' experience in future museums.

3.1.1 Visual Task. The undisclosed purpose of this task was to get participants to encode visual stimuli in their memory, and to expose them to the experimental manipulation. The participant was given a smartphone, and was invited to walk along a path in the corridor of a research building. The path was divided into 16 checkpoints, each marked by a QR code (which we will call *marker*) which the participant had to scan with the smartphone. For photos and details about the physical environment and the path, see Appendix A, subsection A.1. When a marker was scanned, a 3D object (which we will call *stimulus*) was displayed on the phone, either in AR or in non-AR, depending on the condition. For images and details about the stimuli, see Appendix A, subsection A.2.

[3] Pre-registration: osf.io/79myf Supplementary material: osf.io/gscpv

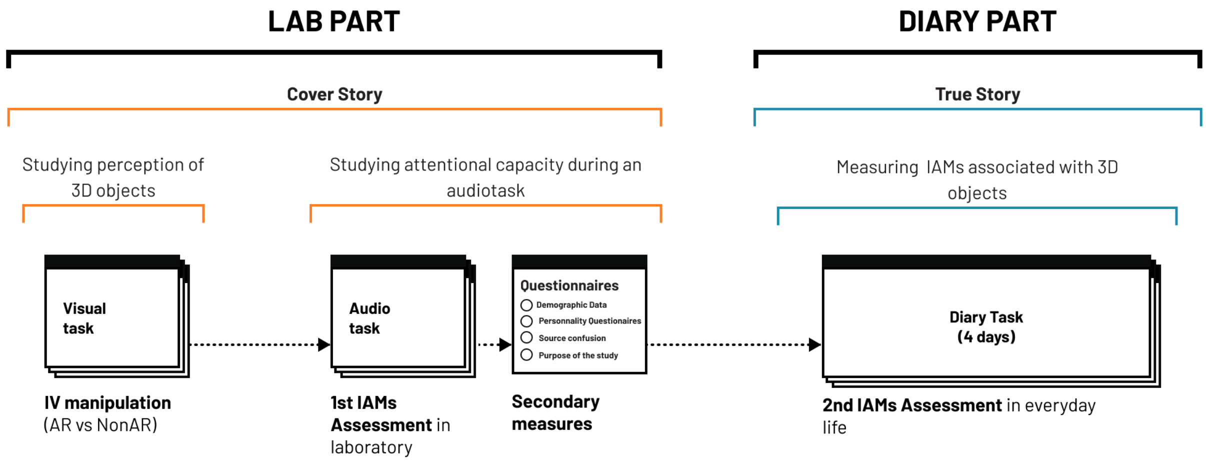


Figure 3: Overview of the study procedure for experiment 1.

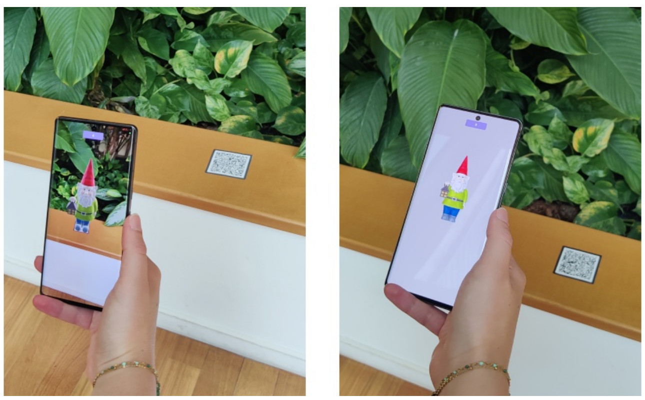


Figure 4: The same stimulus shown in AR (left) and in non-AR (right). In the AR condition, participants were able to move their phone around the virtual object, while in the non-AR condition, they were allowed to rotate the object and zoom with finger gestures.

In the AR condition, the stimulus was displayed using video see-through augmented-reality, in such a way that it appeared to be sitting on the physical surface where the marker was placed (Figure 4-left). The participant could walk around the object. In the non-AR condition, the same stimulus was displayed on a neutral gray background, and could be rotated and zoomed with finger gestures as in a conventional 3D object viewer (Figure 4-right).

The participant was asked to rate each stimulus according to: *i)* its personal relevance (question item inspired by Bellezza and Hoyt [8]), and *ii)* its visual realism (question item inspired by Carlin et al. [21]). The undisclosed purpose of those two questions was merely to draw the participant's attention to the stimuli and facilitate their encoding in memory, by encouraging the participant to process them on a cognitive (first question) and perceptual (second question) level. Each stimulus was displayed for 30 seconds, during which the participant had 10 seconds to answer each of the two questions. The participant previously practiced on dummy stimuli before walking through the path.

One reason why the real purpose of the visual task was hidden to the participant was to ensure that the encoding of the stimuli in memory is *incidental* [57]: if the participant does not expect to have to report their memories of the stimuli later on, they have no reason to try to memorize them. This choice was made to increase ecological validity, as there is rarely deliberate memorization taking place in the real-life situations we want to capture (see subsection 2.1).

3.1.2 Audio Task. Once the visual task was completed and after a short pause, the participant was equipped with headphones and invited to walk through the same path again, although this time, the stimuli were not displayed. They were told the goal now was to measure their auditory concentration, in a task that emulates a museum visit with an audio guide. They were asked to perform three tasks simultaneously:

- *Marker scanning task.* The participant was asked to scan the markers in the same order as before, without hurrying (again, no stimulus was displayed). The undisclosed goal of this task was to re-expose the participant to the physical context where each stimulus was seen. The reasoning is that elements of the physical context (for example, a chair) may have been encoded in memory together with the stimulus (for example, a virtual teapot), and might therefore serve as a cue to an IAM of the stimulus.
- *Auditory vigilance task.* Throughout the entire audio task, the headphones emitted a short sound every 2 to 5 seconds: either a high-pitched beep in the left ear (the *target*), or a low-pitch beep in the right ear (the *distractor*). The participant was asked to keep track of the number of targets by pressing a large button on the smartphone each time they heard one, which incremented a counter on the screen. As we saw in subsection 2.3, the IAM literature suggests that such monotonous tasks promote mind wandering and thus the emergence of IAMs, while occupying participants enough to discourage voluntary recall efforts.
- *Thought reporting task.* The participant was told it was possible and normal for thoughts to surface during the audio task, and that the researchers were interested in learning about them to better

understand their level of concentration. The participant was asked to press a large button on the smartphone each time a thought surfaced. Doing so paused the audio track and displayed a text box for entering a description of the thought (short keywords were allowed). Once done, the audio track resumed.

The participant completed a thorough tutorial before the audio task to ensure that they were comfortable with this triple task.

After the audio task was completed, the participant was invited to sit in front of a laptop showing all the thought descriptions they previously reported, and had the opportunity to edit them, for example to unpack and explain keywords. They were also given a series of short questionnaires, which included a question asking about their familiarity with AR, a question about their familiarity with the physical environment where the experiment took place, and a question meant to verify if they believed in the cover story (details in Appendix A, subsection A.3). Finally, the participant was told the true purpose of the study and given information about the optional follow-up, the diary part. The lab part lasted 45 to 60 minutes in total.

This experimental protocol as well as the instructions were inspired from existing lab protocols for measuring IAMs (e.g., [16, 77], see subsection 2.3). However, previous studies involved participants sitting in front of a computer screen and seeing cues as words presented on the screen. To our knowledge, this is the first protocol that has been adapted to a more ecological situation where participants walk through a physical environment with physical cues.

3.1.3 Diary Task. The optional diary part took place during the participant's regular daily life and and lasted four days (but only required a few minutes of involvement per day). Its purpose was to complement our lab measurements of IAMs with measurements having greater ecological validity, where more time elapses between encoding and recall, and where cues are not controlled.

Participants who volunteered signed a Web-based consent form on their personal phone and the experimenter gave instructions on how to create a phone shortcut to the Web form for reporting IAMs, and on what was expected from them. The participant was instructed to record any IAM of the experiment stimuli as soon as they could, using the Web form. The form first asked them to confirm that the memory was involuntary and was related to one the 3D objects seen in the lab, during the visual task. Participants then had to specify if the cue was in their thoughts or in their physical environment, briefly describe the cue, select the stimulus remembered, rate its vividness on a 5-point scale, and rate their level of concentration when the IAM occurred on a 5-point scale.

3.2 Randomization and Study Design

During the visual task, participants scanned 16 markers, each of which triggered the display of a specific stimulus (e.g., a watering can, a chest,...). All participants saw the same 16 markers and stimuli, in the same order. We created a random mapping (*mapping A*) between stimuli and display conditions ensuring that 8 stimuli are mapped to AR and 8 to non-AR. We then derived *mapping B* that was the exact opposite, i.e., the AR and non-AR conditions were swapped for all stimuli. Participants were randomly assigned to either mapping A or mapping B. The full details of the procedure are in the preregistration and supplementary material.

In summary, the experiment followed a mixed design, with *display type* (AR or non-AR, see again Figure 4) as a within-subject variable, and *mapping* (A or B) as a between-subject variable.

3.3 Apparatus

During the lab phase, all tasks were carried out on a Google Pixel 7 Pro, and participants were equipped with a Jabra Evolve 75 MS headset during the audio task. The mobile applications were developed in C# with Unity (unity.com). We used the Vuforia Augmented Reality SDK (developer.vuforia.com) to implement targets and display AR content. We did extensive pilot testing to ensure that the apps were usable and robust, and that instructions were clear. During the diary phase, the diary task was carried out on the participant's personal mobile phone. The diary app was implemented as a website using HTML and CSS, and a shortcut to the website was added to the home screen of the participants' mobile phone. Each time an IAM was reported, the data was sent to our server using the HTML POST protocol, and a unique participant ID as parameter. Participants were given an explanation of how to remove the shortcut on their phone at the end of the diary period. The source code of all apps are available in our OSF repository.

3.4 Participants

3.4.1 Sample Size and Inclusion Criteria. We decided to recruit enough participants to obtain valid data from 30 participants for the diary part. We estimated that a sample size of 30 would give us a statistical power[4] of 0.5 to 0.9 under reasonable model assumptions.[5] Participants had to be between the ages of 18 and 65, speak French, have no motor or sensory impairment, and not be on medication that impairs cognition.

3.4.2 Data Exclusion Criteria. We decided that data from a participant would be excluded from our analyses if there is evidence that *i)* the participant was not blind to the true purpose of the experiment, as determined by qualitatively examining their answer to the deception check questionnaire (subsection A.3); or *ii)* the participant did not understand the audio vigilance task or did not perform it seriously, as determined by computing task accuracy. In addition, we decided to exclude participant data from the diary part if the participant signed the consent form but did not report a single IAM. Again, the full details are available in our preregistration and in the supplementary material shared on OSF.

3.4.3 Study Sample. We recruited 33 participants for the lab part through local and mail advertising, among whom 30 agreed to continue with the diary part. No participant met our data exclusion criteria for the lab part. For the diary part, due to the high difficulty of finding participants (our institution was unable to offer participant payment at the time), we decided to keep data from participants who signed the consent form but did not report any IAM

[4]Statistical power refers to the probability of detecting an effect of a certain postulated size [6, Chap.8].

[5]We ran Monte Carlo simulations to estimate the power of the Bayesian analysis described in subsection 3.5. Our simulations required assuming a *base rate* of IAMs (average number of IAMs reported per stimulus in the non-AR condition) and a *rate ratio* (multiplicative increase in IAMs in the AR condition). These concepts are detailed in subsection 3.5. Assuming a rate ratio of 2 and a base rate of 0.1 IAMs, the power is 0.9 with 30 participants; for a base rate of only 0.05, it drops to 0.5. The simulation code and results were preregistered and are available on OSF.

(N=17). This change and other minor changes to the preregistration are fully documented in the OSF material.

Of the 33 participants, 16 identified as women and 17 as men, and they were aged between 20 and 47 (M = 28). In terms of their familiarity with the experimental environment, 14 reported being there for the first time, 6 reported having passed by occasionally, 12 reported passing by daily, and 1 participant chose to not answer. In terms of their familiarity with AR, responses were also very diverse: . The most common response was 4 on the 1–7 scale (moderately familiar, reported by 10 participants), and the second most common response was 5 (reported by 8 participants). Meanwhile, 3 participants responded 1 (not familiar at all), and 4 responded 7 (highly familiar).

3.5 Planned Analyses: Methods

3.5.1 Lab Part – Qualitative Coding. To be able to count IAMs in the lab part, we needed to know which reported thoughts referred to an experimental stimulus. This was done through qualitative coding, by two coders who were blind to which experimental condition was assigned to which stimulus. We obtained a Cohen's Kappa of $\kappa = 0.89$, reflecting high agreement. Disagreements were resolved through discussions. The supplementary material includes the codebook, data, and more details.

3.5.2 Lab Part – Bayesian Estimation. We then analyzed the data using hierarchical Bayesian Poisson regression [56]. Poisson regression is a statistical method used to model count data, where the outcome variable represents the number of occurrences of an event (IAMs in our case)[6]; Hierarchical Poisson regression is an extension that accounts for data structured at multiple levels, allowing us to model variability across both participants and stimuli; Finally, the Bayesian version provides probabilistic estimates of effect sizes, offering easy-to-interpret measures of uncertainty [43].

We fitted four Poisson models of different complexities. All models included a global intercept α, and a parameter β for the effect of display type (D). Some models additionally included random intercepts for individual participants (P), and/or random intercepts for different stimuli (S):

model-DPS: $\text{IAMs} \sim \text{Poisson}(\lambda),\ \log(\lambda) = \alpha + \beta \cdot \text{D} + u_\text{P} + u_\text{S}$
model-DP: $\text{IAMs} \sim \text{Poisson}(\lambda),\ \log(\lambda) = \alpha + \beta \cdot \text{D} + u_\text{P}$
model-DS: $\text{IAMs} \sim \text{Poisson}(\lambda),\ \log(\lambda) = \alpha + \beta \cdot \text{D} + u_\text{S}$
model-D: $\text{IAMs} \sim \text{Poisson}(\lambda),\ \log(\lambda) = \alpha + \beta \cdot \text{D}$.

In all four models, β is the effect of display type with a log parameterization. Exponentiating β yields a *rate ratio*, which is the ratio between the expected number of IAMs with AR and the expected number of IAMs with non-AR, for all participants and stimuli. For example, a rate ratio of 2 means that AR yields twice as many expected memories than non-AR. The rate ratio being a easy-to-interpret measure of effect size in our case, it is the model parameter we will estimate to answer our research question.

All four models used the same weakly-informed priors, meaning they incorporate minimal assumptions about the data and allow the analysis to be driven primarily by the observed evidence [33]. The prior for the rate ratio is centered at 1, meaning it reflects an initial assumption of no difference between conditions. Full details of the Bayesian analysis are available in the supplementary material in the form of R Markdown files and Stan models, which have been tested on simulated data and preregistered before the data was collected.

After fitting the four models, we selected the best model using the Leave-One-Out Information Criterion (LOOIC), which is a measure of a model's predictive power [91]. Choosing a model with good predictive power helps reduce the risk of overfitting and increases the likelihood that the estimated effects are robust and reliable. LOOIC estimates a model's predictive power by systematically leaving out each observation and assessing how well the model predicts it from the remaining data [91]. This model selection process was also preregistered.

In what follows, all reported point estimates refer to the median of the posterior distribution, and all reported interval estimates are 95% equal-tailed credible intervals [42].

3.5.3 Diary Part. For the diary part, the data was analyzed in the same way as the lab part, except participants selected the experimental stimuli from a drop-down menu, so no qualitative coding was required. The same Bayesian analysis was used.

3.6 Planned Analyses: Results

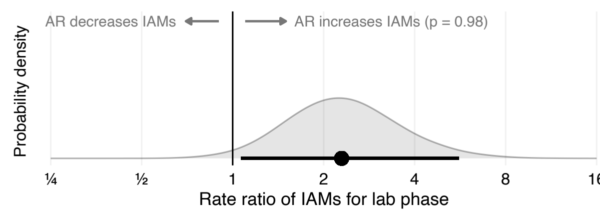


Figure 5: Results for the lab part of experiment 1, showing the point estimate (black dot), 95% credible interval (error bar), and posterior distribution (gray density curve) of the rate ratio on a log scale. The *rate ratio* is the expected number of IAMs for AR divided by the expected number of IAMs for non-AR, overall. For example, a rate ratio of 2 indicates that AR leads to twice as many IAMs overall. The plot also reports the posterior probability that AR leads to more IAMs (gray annotation on top).

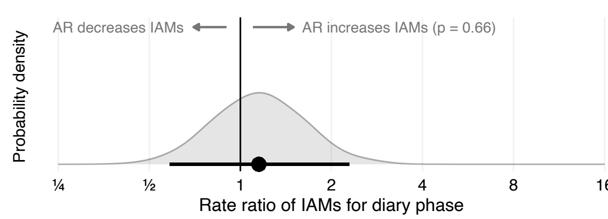


Figure 6: Results for the diary part of experiment 1: point estimate, 95% credible interval, and posterior distribution of the rate ratio on a log scale, with the posterior probability that AR leads to more IAMs.

[6] We modeled the dependent variable as a count because participants were allowed to report IAMs of the same stimulus more than once. Nevertheless, Poisson regression can model rare events, i.e., when the count is most of the time 0, sometimes 1, and never or almost never more than 1.

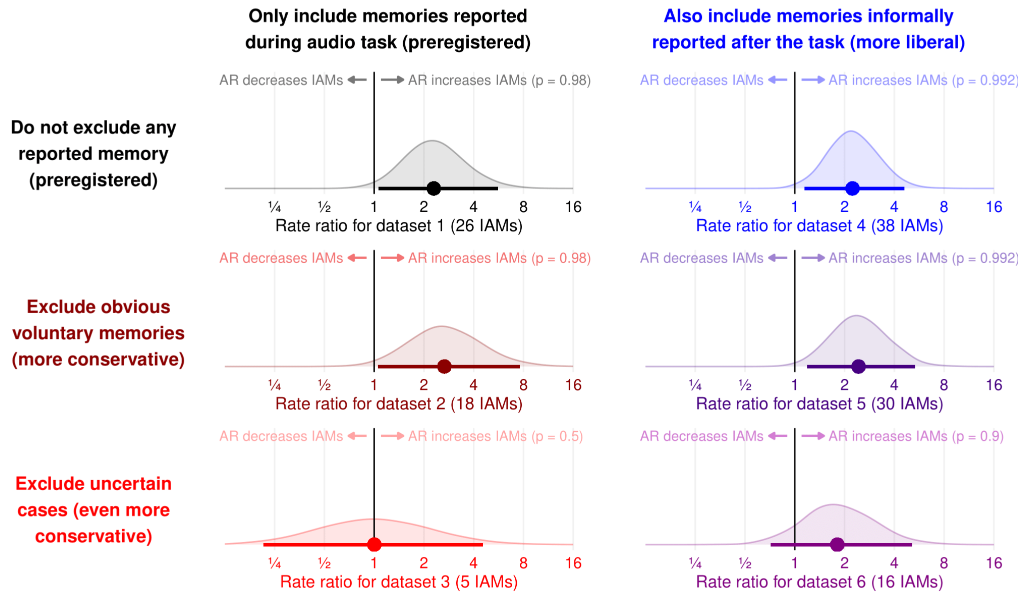


Figure 7: Multiverse analysis of the lab data. Each plot shows the results for a particular experimental dataset, obtained by using different criteria for including or excluding IAMs. The top-left plot shows the results of the preregistered analysis (also in Figure 5).

3.6.1 Lab Part. During the audio task, a total of 224 thoughts were reported, 26 of which were IAMs of experimental stimuli. Of these, 20 were memories of stimuli shown in AR, and 6 were memories of stimuli shown in non-AR. The automatically-selected Bayesian model was *model-DP*. The regression results with this model are plotted in Figure 5. The estimated rate ratio of IAMs is 2.3, 95% CI [1.1, 5.6], and the posterior probability of the rate ratio being greater than one (i.e., AR leads to more IAMs) is $p = 0.98$. Therefore, the lab data provides good evidence that AR elicits more IAMs than non-AR overall, in our lab settings.

3.6.2 Diary Part. During the diary task, a total of 31 IAMs were reported, 17 of which were of stimuli shown in AR, and 14 of stimuli shown in non-AR. The automatically-selected Bayesian model was *model-DPS*. The regression results are plotted in Figure 6. The estimated rate ratio of IAMs is 1.2, 95% CI [0.58, 2.3], and the posterior probability of the rate ratio being greater than one is $p = 0.66$. Therefore, the results for the diary part are inconclusive: the data does not allow us to conclude about the direction of the effect.

3.7 Additional Analyses: Multiverse Analysis

As a post-hoc (non-preregistered) analysis, we carried out a multiverse analysis of our lab data. A *multiverse analysis* involves conducting multiple analyses on the same experimental dataset to assess the extent to which a finding is robust or instead sensitive to arbitrary analytical decisions [86]. We decided to carry out a multiverse analysis for two reasons:

(1) *Some IAMs were reported but not analyzed.* When asked if they forgot to report IAMs of stimuli (see subsection A.3), 6 participants reported extra IAMs, including a participant who stated they thought they were not supposed to report experiment-related thoughts. We wanted to understand if adding those IAMs would change our results.
(2) *Some non-IAMs may have been inadvertently included in the analysis.* After examining the memories reported in the lab data, we identified that 8 of them suggest potential effortful recall. For example, one reported memory stated: *"What was here again? Oh yes, a watering can."*. We labeled them as obvious voluntary autobiographical memories (VAMs). In addition, two participants reported many more memories of stimuli than the rest of the participants (resp. 9 and 11 out of 16 stimuli), implying that some or possibly all of their reported memories may be VAMs. We labeled their memories as uncertain (i.e., either IAM or VAM).

We then ran six versions our previous Bayesian estimation, where we either:

- *(i)* excluded or *(ii)* included informally-reported IAMs,
- *(i)* included all memories of stimuli, *(ii)* excluded obvious VAMs, or *(iii)* excluded both obvious and uncertain VAMs.

All decisions and analyses were made without knowledge of which reported memory was assigned to which experimental condition. The results of the 2×3=6 analyses are reported in Figure 7. The top-left plot is our preregistered analysis (also in Figure 5), while the remaining five are alternative analyses. Overall there is consistent evidence that AR increases the frequency of IAMs, but the evidence is weaker when all uncertain IAMs are excluded (bottom-right), and it is null if we also do not include informally-reported IAMs (bottom-left). The reason is likely insufficient data, i.e., IAMs are too rare to be able to conclude.

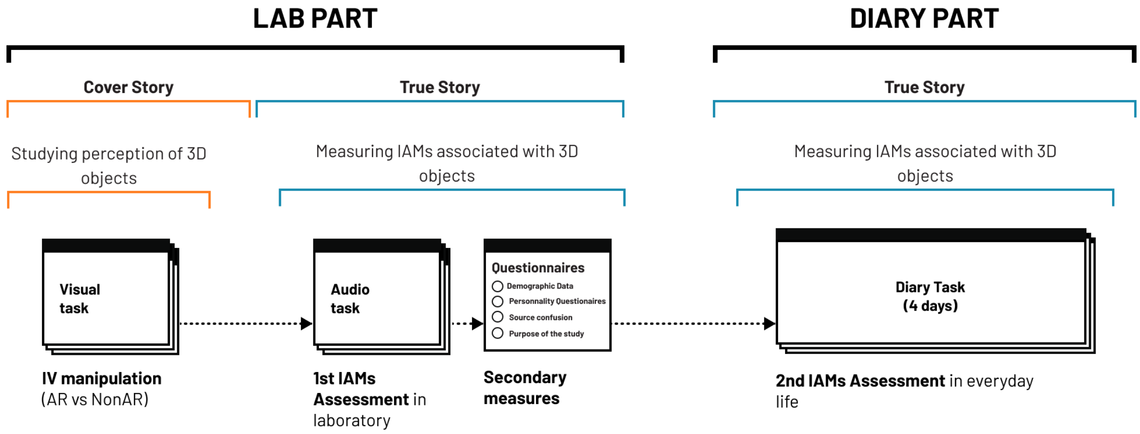


Figure 8: Study procedure for experiment 2 (same as in Figure 3 except for the switch from cover story to true story).

3.8 Discussion

We found evidence that AR produces more IAMs than conventional displays in our lab settings—probably roughly around twice as many. In contrast, data from our diary study is inconclusive—there is almost certainly an effect [23] but a larger sample is needed to measure its direction and magnitude. Despite this inconclusive result and despite the overlap in the interval estimates between the lab and the diary phase (compare Figure 5 and Figure 6), there is some evidence for a weaker effect of AR in the diary phase. We will discuss possible reasons for this later on.

We were surprised by the low number of IAMs reported overall, especially during the lab phase. Overall, each time a participant was exposed to a stimulus, there was only a 5% chance they would subsequently report it as a thought during the audio task. Furthermore, as many as 27 out of 33 participants (82%) did not mention any stimulus in any of their reported thoughts. Some of them likely misunderstood the instructions and thought that recollections of experimental stimuli did not qualify as wandering thoughts that needed to be reported. In fact, this misunderstanding was explicitly mentioned by a participant, and 3 more participants who initially did not report any memory of the stimuli reported some after having been asked. Since people tend to rapidly forget having had IAMs [9, 14], it is likely that more IAMs were actually experienced, but were not reported or did not even cross the participant's awareness threshold because they had no explicit instruction to monitor them [7, 90]. Similarly, the vigilance task may have produced many more wandering thoughts than the participants were willing to record, and in the absence of instructions for filtering them, many IAMs might have been lost. Although we were still able to find evidence for an effect of AR, the relative rarity of the reported IAMs in our methodology hurts statistical power, which can be an issue for experiments with smaller sample sizes and/or effect sizes.

4 Experiment 2

Experiment 2 is a replication of experiment 1, with the difference that we used a method for measuring IAMs in the lab that is meant to maximize the number of reported IAMs: the true purpose of the study was revealed before the audio task, and participants were explicitly asked to report IAMs of the stimuli. In addition, we made several minor improvements to the experiment protocol, including clarifications of the instructions.

The experiment was approved as an amendment by the same ethical review board and was pre-registered on OSF[7].

4.1 Procedure

The procedure is summarized in Figure 8. Like before, the participant was first invited to the lab, and the study's stated purpose was again to study new ways of enhancing visitors' experience in future museums. The reason for the deception was to ensure incidental encoding of the stimuli in memory during the visual task, as in experiment 1.

4.1.1 Visual Task. As before, the participant was asked to follow a path, scan markers, and rate stimuli. The protocol and the instructions were the same as in experiment 1, but due to maintenance and repair works being done in the corridor used in experiment 1, the task took place in a different location (two adjacent experimental rooms in the same building). The markers and stimuli were the same as in experiment 1, but were laid out differently. The full details are in subsection A.1.

4.1.2 Audio Task. In contrast to experiment 1, before the audio task started, the participant was told that the true purpose of the study is to measure IAMs, and they were explained what IAMs are. However, our specific research questions and hypotheses were not mentioned. The audio task was the same as in experiment 1, and consisted of the same three subtasks: a marker scanning task, an auditory vigilance task, and a thought reporting task. However, the instructions for the thought reporting task differed in that the participant was asked to report thoughts related to the stimuli, instead of reporting any wandering thought. If the participant had a voluntary recall of a stimulus, they were asked not to report it unless they were unsure if it was voluntary. The reporting form was adapted accordingly: the participant was asked to report which stimulus was associated with the thought, and to rate the extent to which the thought was involuntary or voluntary, on a scale from 1

[7] Preregistration: osf.io/4u6ex. Supplementary material again available at osf.io/gscpv.

(fully involuntary) to 6 (fully voluntary). They were explained the meaning of the scale prior to the audio task. The participant was also invited to report any additional detail about the memory, such as a cue description or associated thoughts.

The same questionnaires as in experiment 1 were given to the participant after the audio task, except that the deception check questionnaire asked whether the participant had doubts during the visual task specifically, and the question about familiarity with the environment was not included due to an oversight.

4.1.3 Diary Task. The diary task was identical to experiment 1. We kept it to get more data for this task, and to verify if we would observe an effect in the same direction.

4.2 Randomization and Study Design

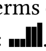
The randomization method and study design were the same as in experiment 1.

4.3 Participants

4.3.1 Sample Size and Inclusion Criteria. Our target was still 30 participants producing valid data and going all the way through the optional diary phase. Inclusion criteria were the same.

4.3.2 Data Exclusion Criteria. We preregistered the same exclusion criteria as in experiment 1 for the lab part (evidence of not being blind to the true purpose of the experiment during the visual task, in order to ensure incidental encoding, or poor performance at the auditory vigilance task). For the diary phase, consistently with our modification to our preregistration of experiment 1, we decided to keep the data from all participants who signed up for that phase, whether or not they reported IAMs.

4.3.3 Study Sample. We recruited new participants, again through local and mail advertising. We obtained valid data from 20 participants for the lab part, and 22 participants for the diary part. Although those numbers are well below our preregistered goal of 30, we had to exclude a significant amount of data from the 30 participants we initially recruited and were unable to find enough volunteers to replace them (again, administrative issues prevented us from offering payment). Data from 4 participants was excluded due to a protocol error that prevented them from completing the deception check questionnaire. Four additional participants reported having doubts about the true purpose of the study during the visual task, and thought it might be about memory, leading us to also exclude their data.[8] Finally, 2 participants performed poorly in the audio task and their data was excluded from the analysis of the lab part, but not of the diary part.

Of the 22 participants whose data was analyzed, 11 identified as women and 11 as men, and they were aged between 18 and 49 years (M = 29). In terms of their familiarity with AR, responses were again very diverse: . The most common response was 6 on the 1–7 scale (just below highly familiar, reported by 6 participants). Meanwhile, 2 participants responded 1 (not familiar at all), and no participant responded 7 (highly familiar).

[8]Comments from those participants suggested that experiment 2 raised more suspicions due to the nature of the physical environment (two experiment rooms, which suggested a study on memory).

4.4 Planned Analyses: Methods

The analysis methods were essentially the same as in experiment 1. For the lab part, the reported thoughts were again assigned to experimental stimuli by two independent coders (we obtained perfect agreement, i.e., $\kappa = 1$). Then, answers to the voluntariness question (which asked to rate the extent to which each thought was involuntary or voluntary on a 1–6 scale) were used to filter out possible VAMs. Specifically, all memories for which people responded 4, 5, or 6 were excluded from the rest of the analyses. Therefore, a memory was considered as an IAM if the participant responded 1, 2, or 3 to the voluntariness question. The IAM data was then analyzed using the same Bayesian analysis procedure as the one detailed in subsection 3.5.

For the diary part, no change was made to the analysis methods.

4.5 Planned Analyses: Results

4.5.1 Lab Part. During the audio task, a total of 141 thoughts were reported, among which 114 were IAMs of a stimulus. This time, each time a participant saw a stimulus, there was a 36% chance they would subsequently report it as an IAM during the audio task, which is considerably more than the 5% of experiment 1. Furthermore, only 4 participants out of 20 reported no IAM (i.e., 20%, vs. 82% in experiment 1).

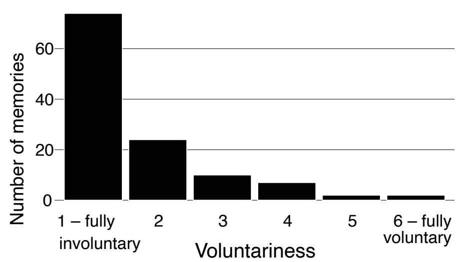


Figure 9: Distribution of responses to the question: "to which extent was this thought voluntary?"

For context, Figure 9 shows the distribution of responses to the voluntariness question for all reported memories (i.e., both VAMs and IAMs). It can be seen that the majority of the reported memories were rated as fully involuntary.

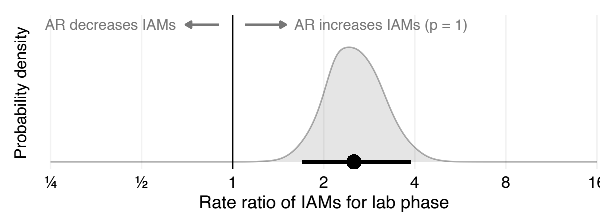


Figure 10: Results for the lab part of experiment 2: point estimate, 95% credible interval, and posterior distribution of the rate ratio on a log scale, with the posterior probability that AR leads to more IAMs.

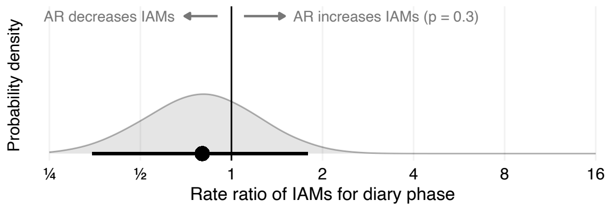


Figure 11: Results for the diary part of experiment 2: point estimate, 95% credible interval, and posterior distribution of the rate ratio on a log scale, with the posterior probability that AR leads to more IAMs.

Of the 114 IAMs reported, 82 were memories of stimuli shown in AR, while 32 were memories of stimuli shown in non-AR. The automatically-selected Bayesian model was *model-DPS*. The regression results with this model are plotted in Figure 10. The estimated rate ratio of IAMs is 2.5, 95% CI [1.7, 3.9], and the posterior probability of the rate ratio being greater than one (i.e., AR leads to more IAMs) is extremely close to 1. Therefore, we have very strong evidence that AR elicits more IAMs than non-AR in our lab settings.

4.5.2 Diary Part. During the diary task, a total of 22 IAMs were reported, 10 of which were of stimuli shown in AR, and 12 of stimuli shown in non-AR. The selected Bayesian model was *model-DPS*. The regression results are plotted in Figure 11. The estimated rate ratio of IAMs is 0.8, 95% CI [0.35, 1.8], and the posterior probability of the rate ratio being greater than one is $p = 0.3$. Therefore, the results for the diary part are again inconclusive: the data does not allow us to conclude about the direction of the effect. In addition, since the point estimate is on the opposite side of experiment 1, combining the data from both experiments will still yield inconclusive results.

4.6 Additional Analyses: Memory Cues

In this section, analyses were conducted post-hoc and not preregistered, and are therefore exploratory. The goal of these extra analyses was to better understand what cues triggered the IAMs, and if there were possible differences between AR and non-AR.

4.6.1 Reporting Time. At what moment were the IAMs reported during the audio task? The two scatterplots in Figure 12 answer this question by showing when each IAM was reported relative to the moment the corresponding marker was scanned, for all IAMs of AR stimuli (left) and for all IAMS of non-AR stimuli (right).

For AR stimuli, many dots lie just above the identity line $y = x$, meaning that many IAMs were reported short after the participant had scanned the marker where the stimulus previously appeared during the visual task. Therefore, it seems that participants often remembered the stimuli right after they were exposed to the place they were shown. Other IAMs were reported way after, way before, or just before the marker was scanned. The tendency to report IAMs shortly after the marker was scanned is also present non-AR stimuli, but is much less pronounced. This is confirmed by looking at the difference in correlations (Pearson's $r = 0.85$ for AR and $r = 0.15$ for non-AR), and at the average time difference between marker scanning and IAM reporting (50 seconds for AR and 108 seconds for non-AR).

We were unable to conduct a similar analysis for experiment 1 due to a data logging issue.

4.6.2 Cue Categories. We examined the different categories of cues that triggered the IAMs, during both the lab part and the diary part, and in both experiments.

For the lab data, we coded[9] the cue descriptions from experiment 2 to classify IAMs according to whether the cue was likely internal (in the participant's thoughts), external (in the environment), or a mix of both (see again subsubsection 2.2.3 for explanations). The results are shown in Figure 13, first column. For experiment 1, as we did not ask participants to describe the cues, all IAMs are marked as "not specified". For experiment 2, many IAMs are also marked as "not specified" as the question was optional. Among the remaining IAMs for which cue descriptions were specified, about one quarter were reported as external, and a smaller proportion were reported as internal or mixed. External cues tended to be reported more often for IAMs of stimuli that were previously shown in AR. Conversely, internal cues tended to be reported more often for stimuli shown in non-AR.

For the diary part, in both experiments, participants were already asked to classify each IAM cue as "in my thoughts" (internal), "in the environment" (external) or "I don't know". The results are reported in the second column of Figure 13. In experiment 1, most IAMs were reported as resulting from internal cues. In experiment 2, however, many more IAMs were reported as resulting from external cues, especially if the stimulus had been shown in AR.

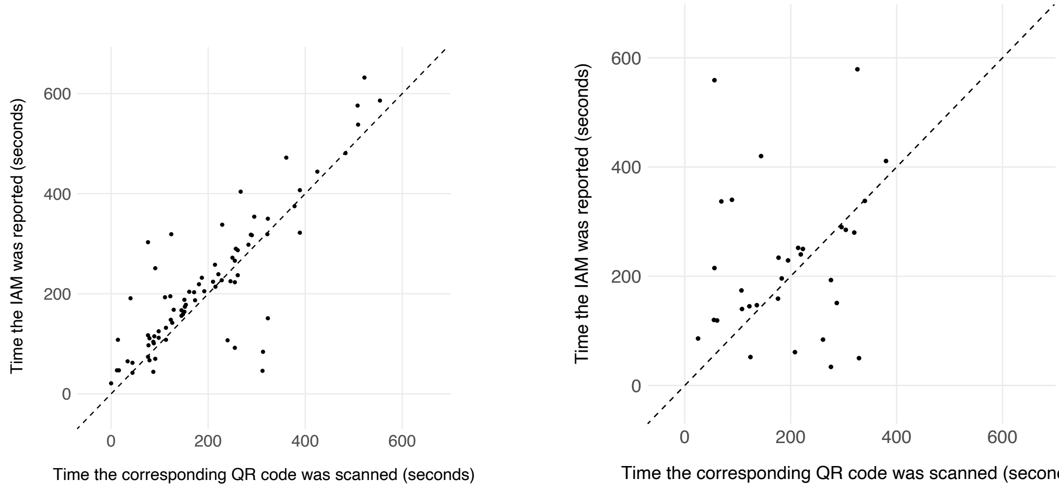
Finally, for the diary part, we used the participants' open-text description of the cues to code whether each cue was related to the experiment (e.g., the participant was thinking about the building where the experiment took place, or saw the diary app on their smartphone). The results are reported in the third column of Figure 13. Overall, slightly more IAMs were triggered by cues related to the experiment. There is no clear difference between AR and non-AR in experiment 1, but in experiment 2, many more IAMs were reported as having been triggered by an experiment-related cue if the stimulus was shown in non-AR.

4.6.3 Newcomers vs. Regular Passersby. In experiment 1, among the participants who did the diary task, 13 of them had reported coming to the experimental environment for the first time (*newcomers*), while 11 of them reported passing by daily (*regular passersby*). Since most regular passersby likely returned to the environment and were re-exposed to the same physical cues, we should expect them to have reported more IAMs during the diary phase. However, newcomers reported an average of 1.7 IAMs per participant (0.9 AR and 0.9 non-AR), while regular passersby reported an average of only 0.7 IAMs per participant (0.4 AR and 0.3 non-AR). Therefore, the data does not support our hypothesis.

4.7 Discussion

Experiment 2 replicates our results from experiment 1. In particular, it confirms that AR yields more IAMs than non-AR in controlled settings, with an effect size estimate consistent with experiment 1 (roughly 2 to 3 times more). In addition, experiment 2 confirms that

[9] All the qualitative coding in this section was done by a single coder, who was blind to which experimental condition corresponded to which IAM.



(a) AR stimuli (b) Non-AR stimuli

Figure 12: IAM reporting times for the lab part of experiment 2, showing the relation between the reported times (y-axis) and the moment the marker was scanned (x-axis). The left scatterplot is for IAMs of AR stimuli, while the right scatterplot is for IAMs of non-AR stimuli.

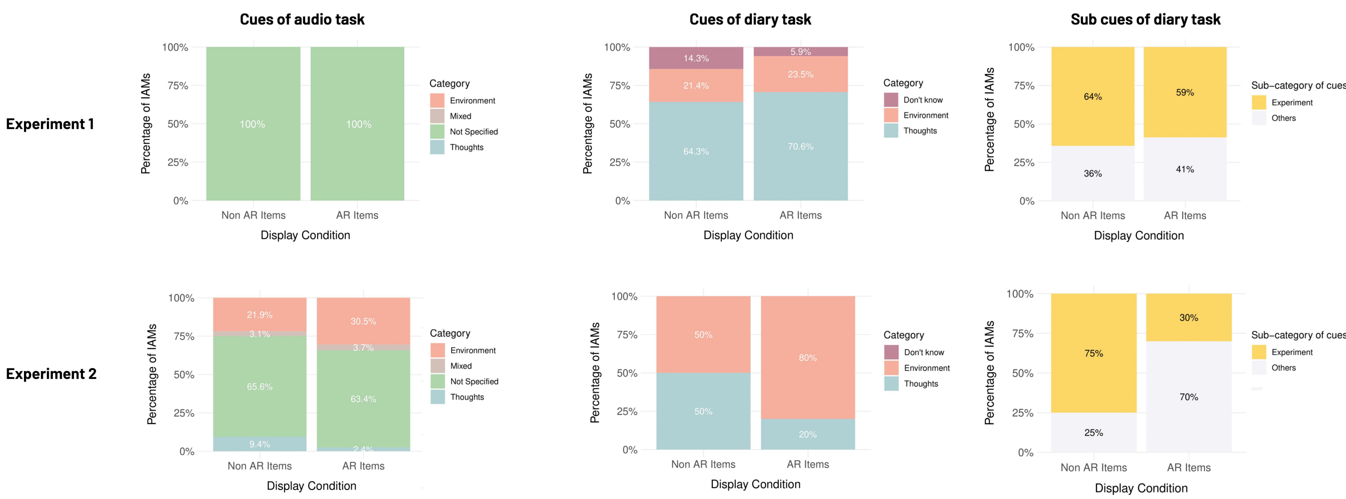


Figure 13: Categories of cues, for both visual and diary parts, and for both experiments.

explicitly prompting IAM monitoring in the instructions results in many more IAMs being reported (roughly 7 times more), which in turn results in a substantially higher statistical power; By comparing Figure 10 with Figure 5, it can be seen that the effect size estimate is much more precise in experiment 2, and the strength of evidence much stronger, despite the lower sample size (i.e., 20 participants vs. 33 participants in experiment 1).

In contrast, our results were again inconclusive for the diary phase. However, this time we have good evidence that the effect of AR on IAMs is lower in the diary phase than in the lab phase—if not in the opposite direction—as there is almost no overlap between the interval estimates in Figure 11 and Figure 10.

5 General Discussion

We review possible explanations for our findings, discuss their potential implications, and suggest directions for future work.

5.1 AR Increases IAMs in Controlled Settings

In the lab part of both experiments, i.e., when participants were subsequently re-exposed to the locations where they saw the stimuli, the stimuli shown in AR led to substantially more IAMs. There can be several reasons for this:

- *AR favors the encoding of physical context*. In section 1 and subsection 2.1, we already speculated that since virtual objects displayed in AR appear embedded in the physical environment, physical objects and features of the environment around the virtual object will tend to be encoded in memory together with the virtual object. Then, when the participant revisits the location where they initially saw the virtual object, the same physical objects and features can act as cues and trigger an IAM of the virtual object. In contrast, in non-AR, virtual objects are completely detached from the physical environment, which will tend to be ignored, since the participant's focus is on the phone. This is consistent with what we know about the associative process on which IAM relies [10, 77], according to which a rich environment, offering multiple potential cues, is beneficial—something that AR can provide through local presence in the physical environment [4, 70]. Both our quantitative and qualitative results are consistent with this explanation. For example, our analysis of cues revealed that IAMs shown in AR tended to be reported soon after the participant revisited the place where the stimuli was shown, while this effect was not as pronounced for stimuli shown in non-AR. Similarly, participants often reported that the cue for their IAMs was external if the stimulus was shown in AR, but less so if it was shown in non-AR.

- *Full-body exploration favors stimulus encoding*. We observed that participants often repositioned their phone (e.g., moving closer, kneeling, or stepping back) after scanning the QR code to better frame the stimulus on their screen. Some continued moving to view the stimulus from different angles. Therefore, participants tended to use their full body to inspect AR objects. In contrast, they only had to use their fingers to inspect non-AR objects. While we do not believe this to be the main explanation for the higher rate of IAMs, it is possible that body movements favored the encoding of the stimuli in memory, thereby amplifying the effect of context encoding. Comments reported by some participants in the lab phase of experiment 2 do suggest that body movements are sometimes encoded with the stimuli, e.g.,: *"I had to move the phone back [when the stimulus appeared]"*.
- *Fun and novelty favor stimulus encoding*. The AR user interface may be more fun to use and more engaging than its non-AR counterpart, which could help create a stronger impression of the AR stimuli in the participant's mind; The effect could be partly due to the novelty of the technology, and may fade once it is not novel anymore [41]. Although fun and novelty could play a role, we again doubt that they can be major contributors to the effect, as they do not explain findings from our cue analysis, and many participants reported being at least moderately familiar with AR technology.

5.2 Experiment 2's Protocol Leads to More IAMs in Controlled Settings

Explicitly instructing participants to report IAMs of experimental stimuli (instead of any wandering thought) led to many more IAMs of both AR and non-AR stimuli, with apparently a similar ratio of IAMs between AR and non-AR. As we already speculated before, this is likely due to two factors:

- *Increasing participants' willingness to report IAMs.* In experiment 1, we had evidence of participants experiencing IAMs but not reporting them. This could be due to a misunderstanding of the experiment instructions, as well as a reluctance to exhaustively report all wandering thoughts. The more precise instructions fixes both issues.
- *Lowering participants' awareness threshold for IAMs.* Selective monitoring is known to lower the threshold needed for IAMs to raise to consciousness [7, 90]. In other terms, having to monitor IAMs tends to cause more IAMs.

5.3 The Effect of AR Fades in the Diary Phase

We were unable to measure an effect of AR in delayed out-of-lab settings, in either of the two experiments. We discuss potential reasons why, which are not mutually exclusive.

- *Cues were encountered less frequently and were less ecphoric.* During the four days of the diary phase, most participants never returned to the location of the experiment and were therefore never re-exposed to the exact physical cues that were encoded with the AR stimuli. Many of them likely saw analogous cues during their daily activities (e.g., a plant, a chair), but due to their dissimilarities with the real cues, their ecphoric power was likely insufficient to trigger IAMs in most cases. Although some participants worked in the building where experiment 1 took place and reported passing through the corridor every day, our data included no evidence that elements of the corridor acted as cues to IAMs during the diary phase. This may be due to the participants' familiarity with the corridor, which may have substantially weakened the cues' ecphoric power [73].
- *Stimuli were forgotten over time.* It is likely that with time, the stimuli became less and less likely to resurface as IAMs, consistent with the phenomenon of exponential decline in information retention as a function of time [29], which is known to also occur for IAMs [39, 85]. Exponential decline means that forgetting is stronger soon after the event, and thus the time before the experiment and the diary phase was likely sufficient for participants to forget most of the stimuli.
- *Participants were forgetful and less inclined to report IAMs.* Participants very likely forgot to report some of the IAMs they experienced. This type of omission is known to occur, especially when smartphones are used as diaries [46]. Similarly, because participants were occupied by their daily activities and were not as focused as in the lab study, they may have been less inclined to report all of their IAMs, especially when it was inconvenient.

5.4 Ecological Validity

As any study, our study needed to strike a balance between experimental control and statistical power on the one hand, and ecological validity on the other hand. In the lab phase, both the visual task and the audio task differ from the way people are exposed to events and to cues in everyday life. In particular, the immediate re-exposure to contextual cues likely greatly increases the likelihood of experiencing an IAM compared to natural exposure to events and cues in real-world settings. Our diary phase, where exposure to cues is no longer controlled, is closer to real-world settings. It is unclear whether more IAMs were experienced in the diary phase than would have been experienced in natural settings. On the one hand, the specific context of stimulus exposure (i.e., a scientific experiment) and the explicit instruction to monitor IAMs could have led to more IAMs than in non-experimental settings. On the other hand, the stimuli used in the experiment were not particularly emotional or self-relevant, and real life likely contains plenty of events that are more striking and more likely to lead to IAMs.

5.5 Possible Future Studies

We believe our study to be the first study looking at the effect of AR on the frequency of IAMs, but many more studies are possible, and follow-up studies are needed to improve our protocol and test the extent to which our findings generalize. Examples include:

- *Other stimuli presentation and encoding approaches:*
  - *Alternative stimuli.* Our collection of stimuli is only a starting point and would benefit from being improved, for example by selecting even more homogeneous stimuli (e.g., we noticed that stimuli like the bust are more memorable than others), and by conducting more rigorous pre-tests. Furthermore, it could be interesting to test other types of stimuli, including stimuli that are inherently more memorable, which could help boost IAMs in the diary phase. This includes emotionally-intense stimuli, as well as stimuli to which participants can personally relate. Personally-relatable stimuli, in particular, are known to be more memorable [26] and could capture real-life situations such as social media browsing or targeted advertising. However, a methodological difficulty is that such stimuli would need to be tailored to participants.
  - *Repeated encoding*. A known solution to forgetting is spaced repetition, i.e., revisiting studied content at multiple time intervals [29, 94]. Such a protocol could be used in IAM studies, e.g., by asking participants to come to the lab and carry out the visual task multiple times. This could have the effect of boosting the frequency of IAMs in the diary phase, and could capture real-life situations where people are repeatedly exposed to the same AR content.
  - *Intentional encoding*. In our experiments, we chose to have participants encode the stimuli incidentally, i.e., without trying to memorize them. However, IAMs resulting from intentional encoding could also be interesting to study, with possible real-life applications in education and behavioral therapy.
  - *Other AR technologies.* We opted for handheld AR in our study for convenience and for greater experimental control, but we are interested in understanding AR generally, without a focus on a particular technology. Nevertheless, it is possible that the type of AR technology used (e.g., handheld vs. head-mounted display) could lead to noteworthy differences. Conducting comparative studies with different AR technologies could give us answers.
  - *Out-of-lab stimuli presentation.* To bring stimulus encoding closer to real-life experiences, participants could be exposed to stimuli on their personal smartphone, during their daily life routine. Alternatively, when special AR equipment needs to be used, experimenters could visit participants at their home to administer the visual task in their own physical environment, or the encoding could take place during public events such as exhibitions.
- *Other cue presentation and recall approaches:*
  - *Delayed re-exposure to cues.* To better disentangle the factors that make AR less effective in the diary phase, a modified protocol could include a second lab phase where participants are invited back to the lab after a few days to complete another audio task.
  - *Longer diary phase.* The more time participants have, the more IAMs they should report, so extending the diary phase may help increase IAMs. However, due to forgetting, this strategy may lead to diminishing returns. It is worth noting that in our study, participants already had considerably more time to experience and report IAMs during the diary task than during the audio task (i.e., 4 days vs. about 10 minutes, a ratio of ×500). This is unsurprising, as the diary phase does not re-expose participants to the exact same contextual cues, and stimuli may be forgotten over time (see subsection 5.3). Yet the extended duration of the diary phase did not appear sufficient to offset these factors.
  - *Use of notifications in diary phase.* As discussed in subsubsection 2.3.1, future studies could use occasional text notifications to keep participants engaged in the diary phase, but random notifications explicitly prompting them to report IAMs (i.e, experience sampling) are likely not adequate [45]. As a more indirect approach, notifications could be used to expose participants to cues that may increase the chances of experiencing an IAM – for example, participants could receive text messages evoking gardening, which may trigger IAMs about the watering can, the garden hose, or the garden gnome. Such notifications could be sent during midday and in the evening, during which mind wandering is more prevalent [82], thereby favoring spontaneous thoughts and recollection [96].
- *Other improvements and research questions:*
  - *Improve measurement*. We framed the voluntariness question in experiment 2 to acknowledge that VAMs and IAMs exist on a continuum [60], but we dichotomized participants' responses for our analysis. Future work could instead employ statistical models that use voluntariness as a continuous predicting variable. Furthermore, current methods for measuring IAMs rely on self-report, which, while useful, comes with inherent limitations such as recall bias, subjectivity, and variability in interpretation. Alternative methods such as those using neuroimaging and monitoring deserve to be explored [35, 36]. Response time could also be used to distinguish between IAMs and VAMs during controlled exposure to cues, as IAMs are recovered more quickly than VAMs [77]. Further research is needed to explore objective measurements that can complement self-reported data and provide a better understanding of the cognitive and emotional processes that underlie technology-mediated IAMs.
  - *Alternative comparisons.* Motivated by the ethical questions and practical scenarios discussed in section 1 and subsection 2.1, our research question focused on comparing AR with conventional computer displays. However, other comparisons could be informative, such as comparing AR objects with physical objects directly placed in the environment. Such studies could capture other types of scenarios than the ones we discussed (e.g., is an AR statue as effective as a physical statue or is AR product preview as effective as a try-before-you buy option in terms of IAM elicitation?) and could help answer new theoretical and methodological questions (e.g., can AR be used as a tool to study human memory, as discussed in subsubsection 2.4.1).

– *Inter-individual differences.* Interpersonal factors (e.g., intellectual humility, mental disorders [67]) and demographic variables (e.g., age [81]) are known to change the accuracy of memory monitoring. Therefore, it seems important to study how inter-individual differences may modulate the effect of AR on IAM.

5.6 Implications of our Findings

In the introduction, we wrote that *"if AR really does promote involuntary memories, the implications can be far-reaching"*. So what are the implications of our findings, now we have ran our study? We empirically demonstrated that the effect exists: in our experimental settings, a visual stimulus shown in AR is roughly twice as likely to resurface as an IAM compared to using a conventional display, if the participant is re-immediately exposed to the physical context of the stimulus. However, we also found that IAMs occur relatively rarely, especially without immediate re-exposure to the context, and that the effect of AR does not seem extraordinarily robust and may fade over time. For the moment, it is not clear that AR can be easily used as a tool to create strong impressions on people's memory; Alternative approaches, such as maximizing the impact of the content itself by appealing to emotions and self-reference—a strategy long known to advertisers and PR specialists—are likely to be more effective. But it is safe to say that AR can be complementary tool that could help amplify these effects.

Importantly, we tested fairly specific conditions, and future studies may uncover much stronger effects. For example, AR technology is getting more and more advanced, and future displays may be able to render virtual content that is indistinguishable from real content [80], which could boost the effectiveness of AR at eliciting IAMs. Furthermore, we discussed other ways the effect of AR could be amplified, such as repeated exposure, which may be feasible in the future on wearable AR displays and could be exploited for abusive advertising. Similarly, re-exposure to novel physical contexts may maximize ecphoric power and thus the effect of AR. To take a real-life example, imagine a maintenance team that needs to intervene on a new site with some safety hazards; The team could be first exposed to site-specific safety instructions using AR technology, which would then resurface as IAMs in the right contexts while the team is working. At the very minimum, our paper demonstrates that the capacity of AR to promote IAMs should be considered as a real possibility, along with other postulated benefits [93] and risks [19, 80] of AR, and should be studied more.

6 Conclusion

We reported the first study showing that AR can elicit more involuntary autobiographical memories (IAMs) than conventional computer displays. We introduced a novel experimental paradigm for investigating the effect of AR on IAMs which combines methodology from psychology and HCI, and we demonstrated three measurement strategies (two based on a lab protocol and one based on a diary protocol) with different trade-offs in terms of internal validity, external validity, and statistical power. We discussed multiple directions for future work, and provide extensive material to encourage other researchers to replicate and extend our work.

Acknowledgments

We thank the members of the Bivwac team for their valuable advice and support throughout this research. We also thank Kan Yao for his help during experiment 2.

References

[1] Amazon.co.uk. 2024. Virtual Try On : Fashion. https://www.amazon.co.uk/b?ie=UTF8&node=29542815031. [Online].
[2] Snap AR. 2023. March 8, 8 Women (video). https://www.youtube.com/watch?v=Ap3fhFFRZP0. [Online; accessed 7-June-2024].
[3] Ambre Assor, Arnaud Prouzeau, Pierre Dragicevic, and Martin Hachet. 2024. Augmented reality waste accumulation visualizations. *ACM Journal of Computing and Sustainable Societies* 2, 2 (2024), 1–29. https://doi.org/10.1145/3636970
[4] Mohammed Thamir Atta, Awanis Romli, and Mazlina Abdul Majid. 2021. The impact of AR/VR on spatial memory performance of learners: A review. In *2021 International Conference on Software Engineering & Computer Systems and 4th International Conference on Computational Science and Information Management (ICSECS-ICOCSIM)*. IEEE, Pekan, Malaysia, 75–79. https://doi.org/10.1109/ICSECS52883.2021.00021
[5] Alan D. Baddeley and Graham Hitch. 1974. *Working Memory.* Elsevier, New York, NY, 47–89. https://doi.org/10.1016/s0079-7421(08)60452-1
[6] Thomas Simon Baguley. 2012. *Serious Stats: A Guide to Advanced Statistics for the Behavioral Sciences.* Palgrave Macmillan, London, UK.
[7] Krystian Barzykowski and Søren Risløv Staugaard. 2015. Does retrieval intentionality really matter? Similarities and differences between involuntary memories and directly and generatively retrieved voluntary memories. *British Journal of Psychology* 107, 3 (Oct. 2015), 519–536. https://doi.org/10.1111/bjop.12160
[8] Francis S. Bellezza and Sandra K. Hoyt. 1992. The Self-Reference Effect and Mental Cueing. *Social Cognition* 10, 1 (March 1992), 51–78. https://doi.org/10.1521/soco.1992.10.1.51
[9] DORTHE BERNTSEN. 1996. Involuntary Autobiographical Memories. *Applied Cognitive Psychology* 10, 5 (Oct. 1996), 435–454. https://doi.org/10.1002/(sici)1099-0720(199610)10:5<435::aid-acp408>3.0.co;2-l
[10] Dorthe Berntsen. 2009. *Involuntary Autobiographical Memories: An Introduction to the Unbidden Past.* Cambridge University Press, Cambridge, UK. https://doi.org/10.1017/cbo9780511575921
[11] Dorthe Berntsen. 2020. Involuntary autobiographical memories and their relation to other forms of spontaneous thoughts. *Philosophical Transactions of the Royal Society B: Biological Sciences* 376, 1817 (Dec. 2020), 20190693. https://doi.org/10.1098/rstb.2019.0693
[12] Dorthe Berntsen. 2021. Involuntary autobiographical memories and their relation to other forms of spontaneous thoughts. *Philosophical Transactions of the Royal Society B: Biological Sciences* 376, 1817 (2021), 20190693. https://doi.org/10.1098/rstb.2019.0693
[13] Dorthe Berntsen and Nicoline Marie Hall. 2004. The episodic nature of involuntary autobiographical memories. *Memory & Cognition* 32, 5 (July 2004), 789–803. https://doi.org/10.3758/bf03195869
[14] Dorthe Berntsen and David C Rubin. 2002. Emotionally charged autobiographical memories across the life span: The recall of happy, sad, traumatic and involuntary memories. *Psychology and Aging* 17, 4 (2002), 636–652. https://doi.org/10.1037//0882-7974.17.4.636
[15] Dorthe Berntsen, David C Rubin, and Sinue Salgado. 2015. The frequency of involuntary autobiographical memories and future thoughts in relation to daydreaming, emotional distress, and age. *Consciousness and Cognition* 36 (2015), 352–372. https://doi.org/10.1016/j.concog.2015.07.007
[16] Dorthe Berntsen, Søren Risløv Staugaard, and Louise Maria Torp Sørensen. 2013. Why am I remembering this now? Predicting the occurrence of involuntary (spontaneous) episodic memories. *Journal of Experimental Psychology: General* 142, 2 (2013), 426–444. https://doi.org/10.1037/a0029128
[17] Simon E. Blackwell, Daniela Dooley, Felix Würtz, Marcella L. Woud, and Jürgen Margraf. 2020. Inducing positive involuntary mental imagery in everyday life: an experimental investigation. *Memory* 28, 9 (Sept. 2020), 1157–1172. https://doi.org/10.1080/09658211.2020.1822413
[18] Elise Bonnail, Julian Frommel, Eric Lecolinet, Samuel Huron, and Jan Gugenheimer. 2024. Was it Real or Virtual? Confirming the Occurrence and Explaining Causes of Memory Source Confusion between Reality and Virtual Reality. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24, Vol. 4)*. ACM, Honolulu, HI, USA, 1–17. https://doi.org/10.1145/3613904.3641992
[19] Elise Bonnail, Wen-Jie Tseng, Mark Mcgill, Eric Lecolinet, Samuel Huron, and Jan Gugenheimer. 2023. Memory Manipulations in Extended Reality. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23, Vol. 4)*. ACM, Hamburg, Germany, 1–20. https://doi.org/10.1145/3544548.3580988
[20] Simon Bowen and Daniela Petrelli. 2011. Remembering today tomorrow: Exploring the human-centred design of digital mementos. *International Journal of Human-Computer Studies* 69, 5 (May 2011), 324–337. https://doi.org/10.1016/j.

ijhcs.2010.12.005
[21] Albert S. Carlin, Hunter G. Hoffman, and Suzanne Weghorst. 1997. Virtual reality and tactile augmentation in the treatment of spider phobia: a case report. *Behaviour Research and Therapy* 35, 2 (Feb. 1997), 153–158. https://doi.org/10.1016/s0005-7967(96)00085-x
[22] Chen Chen, Lei Zhang, Tony Luczak, Eboni Smith, and Reuben F. Burch. 2019. Using Microsoft HoloLens to improve memory recall in anatomy and physiology: A pilot study to examine the efficacy of using augmented reality in education. *Journal of Educational Technology Development and Exchange* 12, 1 (aug 2019), 17–31. https://doi.org/10.18785/jetde.1201.02
[23] Jacob Cohen. 1994. The earth is round (p<. 05). *American psychologist* 49, 12 (1994), 997. https://doi.org/10.1037/0003-066X.49.12.997
[24] Scott N Cole, Søren R Staugaard, and Dorthe Berntsen. 2016. Inducing involuntary and voluntary mental time travel using a laboratory paradigm. *Memory & Cognition* 44, 3 (2016), 376–389. https://doi.org/10.3758/s13421-015-0564-9
[25] M.A. Conway and H.L. Williams. 2008. *Autobiographical Memory*. Elsevier, Amsterdam, Netherlands, 893–909. https://doi.org/10.1016/b978-012370509-9.00135-2
[26] Martin A Conway. 2005. Memory and the self. *Journal of memory and language* 53, 4 (2005), 594–628. https://doi.org/10.1016/j.jml.2005.08.005
[27] Pierre Dragicevic. 2022. Towards immersive humanitarian visualizations. https://doi.org/10.48550/arXiv.2204.01313
[28] Phebe Petra Driebergen. 2019. Validation of the "Mind-Wandering Spontaneous (MW-S) and the Mind-Wandering Deliberate (MW-D)" scales. Individual Differences in Spontaneous and Deliberate Mind-wandering and the Tendency to Experience Positive and Negative Affect. Available at https://archive-ouverte.unige.ch/unige:124843.
[29] Hermann Ebbinghaus. 1913. *Memory: A contribution to experimental psychology.* Teachers College Press, New York.
[30] Ajoy S Fernandes, Ranxiao Frances Wang, and Daniel J Simons. 2015. Remembering the physical as virtual: Source confusion and physical interaction in augmented reality. In *Proceedings of the ACM SIGGRAPH Symposium on Applied Perception*. ACM, Tübingen, Germany, 127–130. https://doi.org/10.1145/2804408.2804423
[31] Georgia A. Floridou and Daniel Müllensiefen. 2015. Environmental and mental conditions predicting the experience of involuntary musical imagery: An experience sampling method study. *Consciousness and Cognition* 33 (May 2015), 472–486. https://doi.org/10.1016/j.concog.2015.02.012
[32] Robert Gardner and Giorgio Ascoli. 2015. The Natural Frequency of Human Prospective Memory Increases With Age. *Psychology and aging* 30 (03 2015), 209–219. https://doi.org/10.1037/a0038876
[33] Andrew Gelman, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari, and Donald B Rubin. 2013. *Bayesian Data Analysis*. Chapman and Hall/CRC, Boca Raton, FL.
[34] Peter Graf and Daniel L Schacter. 1985. Implicit and explicit memory for new associations in normal and amnesic subjects. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 11, 3 (1985), 501–518. https://doi.org/10.1037//0278-7393.11.3.501
[35] N HALL, A GJEDDE, and R KUPERS. 2008. Neural mechanisms of voluntary and involuntary recall: A PET study. *Behavioural Brain Research* 186, 2 (Jan. 2008), 261–272. https://doi.org/10.1016/j.bbr.2007.08.026
[36] Shana A. Hall, David C. Rubin, Amanda Miles, Simon W. Davis, Erik A. Wing, Roberto Cabeza, and Dorthe Berntsen. 2014. The Neural Basis of Involuntary Episodic Memories. *Journal of Cognitive Neuroscience* 26, 10 (Oct. 2014), 2385–2399. https://doi.org/10.1162/jocn_a_00633
[37] Hunter G. Hoffman, Azucena Garcia-Palacios, Ayanna K. Thomas, and Anne Schmidt. 2001. Virtual Reality Monitoring: Phenomenal Characteristics of Real, Virtual, and False Memories. *CyberPsychology & Behavior* 4, 5 (Oct. 2001), 565–572. https://doi.org/10.1089/109493101753235151
[38] Samuel Jacob, Mandar Warde, and Pratibha Dumane. 2020. Impact of Augmented Reality as an ICT tool to Deliver Engineering Education Content. In *2020 International Conference on Convergence to Digital World - Quo Vadis (ICCDW)*. IEEE, Mumbai, India, 1–6. https://doi.org/10.1109/iccdw45521.2020.9318709
[39] Shunji Kamiya. 2013. Relationship between frequency of involuntary autobiographical memories and cognitive failure. *Memory* 22, 7 (Oct. 2013), 839–851. https://doi.org/10.1080/09658211.2013.838630
[40] Joanna Kisker, Thomas Gruber, and Benjamin Schöne. 2019. Experiences in virtual reality entail different processes of retrieval as opposed to conventional laboratory settings: A study on human memory. *Current Psychology* 40, 7 (April 2019), 3190–3197. https://doi.org/10.1007/s12144-019-00257-2
[41] Michael Koch, Kai von Luck, Jan Schwarzer, and Susanne Draheim. 2018. The Novelty Effect in Large Display Deployments – Experiences and Lessons-Learned for Evaluating Prototypes. In *Proceedings of the 16th European Conference on Computer-Supported Cooperative Work - Exploratory Papers*. European Society for Socially Embedded Technologies (EUSSET), Nancy, France. https://doi.org/10.18420/ecscw2018_3
[42] John Kruschke. 2014. *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan.* Academic Press, Amsterdam, Netherlands.
[43] John K Kruschke and Torrin M Liddell. 2018. The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic bulletin & review* 25 (2018), 178–206. https://doi.org/10.3758/s13423-016-1221-4
[44] Lia Kvavilashvili and Simone Schlagman. 2011. Involuntary autobiographical memories in dysphoric mood: A laboratory study. *Memory* 19, 4 (2011), 331–345. https://doi.org/10.1080/09658211.2011.568495
[45] Andrew Laughland. 2017. *Methodological Issues of Quantifying Everyday Memory Phenomena With Paper and Electronic Diaries*. Ph. D. Dissertation. University of Hertfordshire. https://doi.org/10.18745/TH.18407
[46] Andrew Laughland and Lia Kvavilashvili. 2018. Should participants be left to their own devices? Comparing paper and smartphone diaries in psychological research. *Journal of Applied Research in Memory and Cognition* 7, 4 (2018), 552–563. https://doi.org/10.1037/h0101828
[47] Mark R. Leary, Kate J. Diebels, Erin K. Davisson, Katrina P. Jongman-Sereno, Jennifer C. Isherwood, Kaitlin T. Raimi, Samantha A. Deffler, and Rick H. Hoyle. 2017. Cognitive and Interpersonal Features of Intellectual Humility. *Personality and Social Psychology Bulletin* 43, 6 (March 2017), 793–813. https://doi.org/10.1177/0146167217697695
[48] Alexandra S. Levine. 2022. Ukraine's Promotion Of Fake Paris Bombing Video Highlights Risks Of Misinformation. https://www.forbes.com/sites/alexandralevine/2022/03/17/ukraines-promotion-of-fake-paris-bombing-video-highlights-risks-of-misinformation/. [Online; accessed 14-August-2024].
[49] Kenneth Y. T. Lim and Ryan Lim. 2020. Semiotics, memory and augmented reality: History education with learner-generated augmentation. *British Journal of Educational Technology* 51, 3 (Jan. 2020), 673–691. https://doi.org/10.1111/bjet.12904
[50] John Mace. 2005. Priming involuntary autobiographical memories. *Memory* 13, 8 (Nov. 2005), 874–884. https://doi.org/10.1080/09658210444000485
[51] J Mace. 2006. Episodic remembering creates access to involuntary conscious memory: Demonstrating involuntary recall on a voluntary recall task. *Memory* 14 (2006), 917–924. https://doi.org/10.1080/09658210600759766
[52] John H Mace. 2004. Involuntary autobiographical memories are highly dependent on abstract cuing: the Proustian view is incorrect. *Applied Cognitive Psychology* 18, 7 (2004), 893–899. https://doi.org/10.1002/acp.1020
[53] Loretta S. Malta, Cezar Giosan, Lauren E. Szkodny, Margaret M. Altemus, Albert A. Rizzo, David A. Silbersweig, and JoAnn Difede. 2020. Predictors of involuntary and voluntary emotional episodic memories of virtual reality scenarios in Veterans with and without PTSD. *Memory* 28, 6 (May 2020), 724–740. https://doi.org/10.1080/09658211.2020.1770289
[54] Keiichi Matsuda. 2023. Hyper-Reality. http://hyper-reality.co/. [Online; accessed 7-June-2024].
[55] Giuliana Mazzoni, Manila Vannucci, and Iram Batool. 2014. Manipulating cues in involuntary autobiographical memory: Verbal cues are more effective than pictorial cues. *Memory and Cognition* 42, 7 (May 2014), 1076–1085. https://doi.org/10.3758/s13421-014-0420-3
[56] Richard McElreath. 2018. Counting and Classification. In *Statistical rethinking: A Bayesian course with examples in R and Stan*. Chapman and Hall/CRC, Boca Raton, FL, USA, Chapter 10, 291–330.
[57] Barry McLaughlin. 1965. "Intentional" and "incidental" learning in human subjects: The role of instructions to learn and motivation. *Psychological Bulletin* 63, 5 (May 1965), 359–376. https://doi.org/10.1037/h0021759
[58] Scott D Moffat, Alan B Zonderman, and Susan M Resnick. 2001. Age differences in spatial memory in a virtual environment navigation task. *Neurobiology of Aging* 22, 5 (Sept. 2001), 787–796. https://doi.org/10.1016/s0197-4580(01)00251-2
[59] Ine Mols, Elise van den Hoven, and Berry Eggen. 2014. Making memories: a cultural probe study into the remembering of everyday life. In *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational (NordiCHI '14)*. ACM, Helsinki, Finland, 256–265. https://doi.org/10.1145/2639189.2639209
[60] Chris J. A. Moulin, Fabien Carreras, and Krystian Barzykowski. 2022. The phenomenology of autobiographical retrieval. *WIREs Cognitive Science* 14, 3 (Dec. 2022), 1–21. https://doi.org/10.1002/wcs.1638
[61] Francisco Munoz-Montoya, M-Carmen Juan, Magdalena Mendez-Lopez, and Camino Fidalgo. 2019. Augmented reality based on SLAM to assess spatial short-term memory. *IEEE Access* 7 (2019), 2453–2466. https://doi.org/10.1109/ACCESS.2018.2886627
[62] Inter IKEA newsroom. 2017. IKEA Place app launched to help people virtually place furniture at home. https://www.ikea.com/global/en/newsroom/innovation/ikea-launches-ikea-place-a-new-app-that-allows-people-to-virtually-place-furniture-in-their-home-170912/. [Online; accessed 7-June-2024].
[63] Niels Peter Nielsen, Sinué Salgado, and Dorthe Berntsen. 2020. Using virtual reality to examine emotional hotspots and intrusions in the trauma film paradigm. *Journal of Applied Research in Memory and Cognition* 9, 3 (Sept. 2020), 370–380. https://doi.org/10.1037/h0101853
[64] Charlotte R. Pennington, Derek Heim, Andrew R. Levy, and Derek T. Larkin. 2016. Twenty Years of Stereotype Threat Research: A Review of Psychological Mediators. *PLOS ONE* 11, 1 (Jan. 2016), e0146487. https://doi.org/10.1371/journal.

pone.0146487
[65] Daniela Petrelli, Steve Whittaker, and Jens Brockmeier. 2008. AutoTopography: what can physical mementos tell us about digital memories?. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08)*. ACM, Florence, Italy, 53–62. https://doi.org/10.1145/1357054.1357065
[66] Gaën Plancher, Serge Nicolas, and Pascale Piolino. 2008. Virtual Reality as a Tool for Assessing Episodic Memory. In *Proceedings of the 2008 ACM Symposium on Virtual Reality Software and Technology (VRST '08)*. ACM, Bordeaux, France, 167–170. https://doi.org/10.1145/1450579.1450617
[67] Tenelle Porter, Abdo Elnakouri, Ethan A. Meyers, Takuya Shibayama, Eranda Jayawickreme, and Igor Grossmann. 2022. Predictors and consequences of intellectual humility. *Nature Reviews Psychology* 1, 9 (June 2022), 524–536. https://doi.org/10.1038/s44159-022-00081-9
[68] Anne S. Rasmussen and Dorthe Berntsen. 2011. The unpredictable past: Spontaneous autobiographical memories outnumber autobiographical memories retrieved strategically. *Consciousness and Cognition* 20, 4 (Dec. 2011), 1842–1846. https://doi.org/10.1016/j.concog.2011.07.010
[69] Anne S. Rasmussen, Stine B. Ramsgaard, and Dorthe Berntsen. 2015. Frequency and functions of involuntary and voluntary autobiographical memories across the day. *Psychology of Consciousness: Theory, Research, and Practice* 2, 2 (June 2015), 185–205. https://doi.org/10.1037/cns0000042
[70] Philipp A. Rauschnabel, Reto Felix, Chris Hinsch, Hamza Shahab, and Florian Alt. 2022. What is XR? Towards a Framework for Augmented and Virtual Reality. *Computers in Human Behavior* 133 (Aug. 2022), 107289. https://doi.org/10.1016/j.chb.2022.107289
[71] Nicco Reggente. 2023. VR for cognition and memory. *Curr. Top. Behav. Neurosci.* 65 (2023), 189–232. https://doi.org/10.1007/7854_2023_425
[72] Louis Renoult, Muireann Irish, Morris Moscovitch, and Michael D. Rugg. 2019. From Knowing to Remembering: The Semantic–Episodic Distinction. *Trends in Cognitive Sciences* 23, 12 (Dec. 2019), 1041–1057. https://doi.org/10.1016/j.tics.2019.09.008
[73] Jessica Robin, Luisa Garzon, and Morris Moscovitch. 2019. Spontaneous memory retrieval varies based on familiarity with a spatial context. *Cognition* 190 (Sept. 2019), 81–92. https://doi.org/10.1016/j.cognition.2019.04.018
[74] Oscar Rosello, Marc Exposito, and Pattie Maes. 2016. NeverMind: Using Augmented Reality for Memorization. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology (UIST '16)*. ACM, Tokyo, Japan, 215–216. https://doi.org/10.1145/2984751.2984776
[75] Marius Rubo, Nadine Messerli, and Simone Munsch. 2021. The human source memory system struggles to distinguish virtual reality and reality. *Computers in Human Behavior Reports* 4 (Aug. 2021), 100111. https://doi.org/10.1016/j.chbr.2021.100111
[76] Hélène Sauzéon, Bernard N'Kaoua, Prashant Arvind Pala, Mathieu Taillade, Sophie Auriacombe, and Pascal Guitton. 2014. Everyday-like memory for objects in ageing and Alzheimer's disease assessed in a visually complex environment: The role of executive functioning and episodic memory. *Journal of Neuropsychology* 10, 1 (Oct. 2014), 33–58. https://doi.org/10.1111/jnp.12055
[77] S Schlagman and L Kvavilashvili. 2008. Involuntary autobiographical memories in and outside the laboratory: How different are they from voluntary autobiographical memories? *Memory & Cognition* 36, 5 (2008), 920–932. https://doi.org/10.3758/mc.36.5.920
[78] Benjamin Schöne, Marlene Wessels, and Thomas Gruber. 2017. Experiences in Virtual Reality: a Window to Autobiographical Memory. *Current Psychology* 38, 3 (July 2017), 715–719. https://doi.org/10.1007/s12144-017-9648-y
[79] Daniel J. Simons, Ranxiao Frances Wang, and David Roddenberry. 2002. Object recognition is mediated by extraretinal information. *Perception & Psychophysics* 64, 4 (May 2002), 521–530. https://doi.org/10.3758/bf03194723
[80] Mel Slater, Cristina Gonzalez-Liencres, Patrick Haggard, Charlotte Vinkers, Rebecca Gregory-Clarke, Steve Jelley, Zillah Watson, Graham Breen, Raz Schwarz, William Steptoe, Dalila Szostak, Shivashankar Halan, Deborah Fox, and Jeremy Silver. 2020. The Ethics of Realism in Virtual and Augmented Reality. *Frontiers in Virtual Reality* 1 (March 2020), 1. https://doi.org/10.3389/frvir.2020.00001
[81] Julia Sluzenski, Nora Newcombe, and Wendy Ottinger. 2004. Changes in reality monitoring and episodic memory in early childhood. *Developmental Science* 7, 2 (March 2004), 225–245. https://doi.org/10.1111/j.1467-7687.2004.00341.x
[82] Gabriel King Smith, Caitlin Mills, Alexandra Paxton, and Kalina Christoff. 2018. Mind-wandering rates fluctuate across the day: evidence from an experience-sampling study. *Cognitive Research: Principles and Implications* 3, 1 (Dec. 2018), 1. https://doi.org/10.1186/s41235-018-0141-4
[83] S Adam Smith. 2019. Virtual reality in episodic memory research: A review. *Psychonomic bulletin & review* 26 (2019), 1213–1237. https://doi.org/10.3758/s13423-019-01605-w
[84] Peter Sommerauer and Oliver Mueller. 2018. Augmented reality in informal learning environments: Investigating short-term and long-term effects. In *Proceedings of the 51st Hawaii International Conference on System Sciences*. Hawaii International Conference on System Sciences, Hawaii, USA. https://doi.org/10.24251/HICSS.2018.176
[85] Søren Risløv Staugaard and Dorthe Berntsen. 2019. Retrieval intentionality and forgetting: How retention time and cue distinctiveness affect involuntary and voluntary retrieval of episodic memories. *Memory and Cognition* 47, 5 (2019), 893–905. https://doi.org/10.3758/s13421-019-00904-w
[86] Sara Steegen, Francis Tuerlinckx, Andrew Gelman, and Wolf Vanpaemel. 2016. Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science* 11, 5 (2016), 702–712. https://doi.org/10.1177/1745691616658637
[87] E Tulving. 1983. Ecphoric processes in episodic memory. *Philosophical Transactions of the Royal Society B: Biological Sciences* 302 (1983), 361–371. https://doi.org/10.1098/rstb.1983.0060
[88] Endel Tulving. 1985. How many memory systems are there? *American Psychologist* 40, 4 (April 1985), 385–398. https://doi.org/10.1037/0003-066x.40.4.385
[89] Doménique van Gennip, Elise van den Hoven, and Panos Markopoulos. 2015. Things That Make Us Reminisce: Everyday Memory Cues as Opportunities for Interaction Design. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, Seoul, Republic of Korea, 3443–3452. https://doi.org/10.1145/2702123.2702460
[90] Manila Vannucci, Iram Batool, Claudia Pelagatti, and Giuliana Mazzoni. 2014. Modifying the Frequency and Characteristics of Involuntary Autobiographical Memories. *PLoS ONE* 9, 4 (April 2014), e89582. https://doi.org/10.1371/journal.pone.0089582
[91] Aki Vehtari, Andrew Gelman, and Jonah Gabry. 2017. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and computing* 27 (2017), 1413–1432. https://doi.org/10.1007/s11222-016-9696-4
[92] Olga C. Watkins and Michael J. Watkins. 1975. Buildup of proactive inhibition as a cue-overload effect. *Journal of Experimental Psychology: Human Learning and Memory* 1, 4 (July 1975), 442–452. https://doi.org/10.1037/0278-7393.1.4.442
[93] Wesley Willett, Bon Adriel Aseniero, Sheelagh Carpendale, Pierre Dragicevic, Yvonne Jansen, Lora Oehlberg, and Petra Isenberg. 2021. Perception! immersion! empowerment! superpowers as inspiration for visualization. *IEEE Transactions on Visualization and Computer Graphics* 28, 1 (2021), 22–32. https://doi.org/10.1109/TVCG.2021.3114844
[94] Yael Wollstein and Noel Jabbour. 2022. Spaced Effect Learning and Blunting the Forgetfulness Curve. *Ear, Nose & Throat Journal* 101, 9 (Nov. 2022), 42S–46S. https://doi.org/10.1177/01455613231163726
[95] Yuki Yamada, Keisuke Irie, Kota Gushima, Fumiko Ishizawa, Mohammed Al Sada, and Tatsuo Nakajima. 2017. HoloMoL: human memory augmentation with mixed-reality technologies. In *Proceedings of the 21st International Academic Mindtrek Conference (AcademicMindtrek'17)*. ACM, Tampere, Finland, 235–238. https://doi.org/10.1145/3131085.3131097
[96] Ryan C. Yeung and Myra A. Fernandes. 2024. Recurrent involuntary memories and mind wandering are related but distinct. *Psychological Research* 88, 5 (April 2024), 1483–1498. https://doi.org/10.1007/s00426-024-01961-w
[97] Zheng You Lim, Chean Khim Toa, Edwin Rao, and Kok Swee Sim. 2023. Development of Augmented Reality Based Applications for Brain Memory Training. *International Journal on Robotics, Automation and Sciences* 5, 1 (April 2023), 13–20. https://doi.org/10.33093/ijoras.2023.5.1.3

A Additional Experiment Design Info

In this appendix, we describe the physical environments where the two experiments were conducted, the stimuli we used as well as our stimulus selection procedure, and our post-task questionnaires.

A.1 Physical Environments

Experiment 1 took place in the corridor of a research building (the Inria BSO building in Bordeaux, France). We placed 16 paper-based QR codes at different places in the environment (e.g., on a chair, on a wall, on the floor), and they were ordered to form a path. Each QR code triggered the display of the same stimulus (3D object), in either AR or non-AR (Figure 4). Figure 14 shows a floor plan of the experimental environment, and the sequence of QR codes arranged from right to left. For each QR code, the figure shows a photo of its immediate physical surroundings as well as the name of its associated stimulus (stimuli are shown in Figure 16 and discussed in more detail in the next subsection).

Experiment 2 took place in two adjacent experimental rooms in the same research building. A new path was created with the same 16 QR codes and stimuli, but placed at different locations and in a different order (see Figure 15).

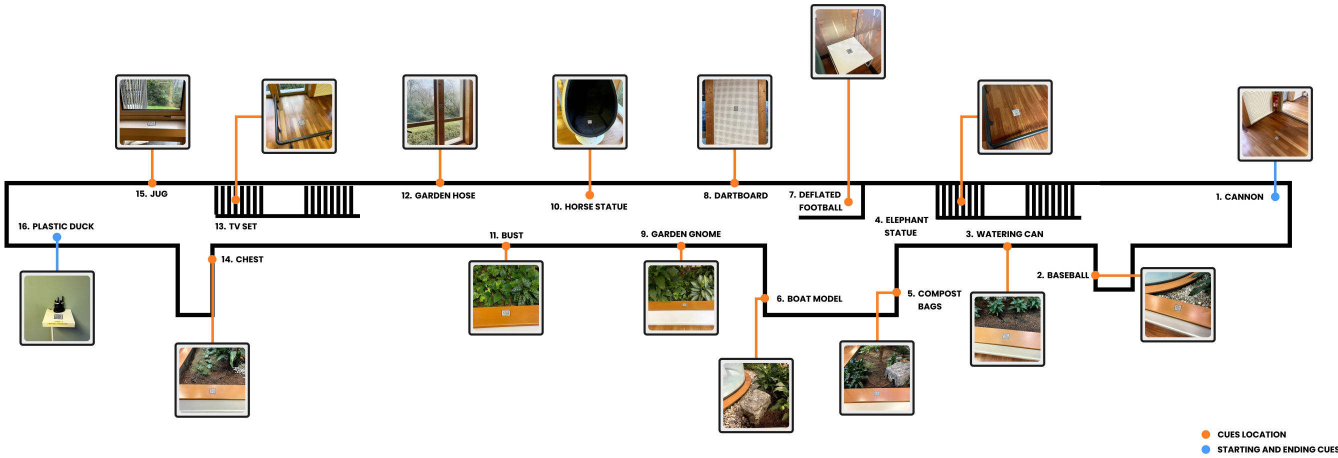


Figure 14: Experiment 1: overview of the physical environment where the QR codes were placed, and the path took by the participants (from right to left).

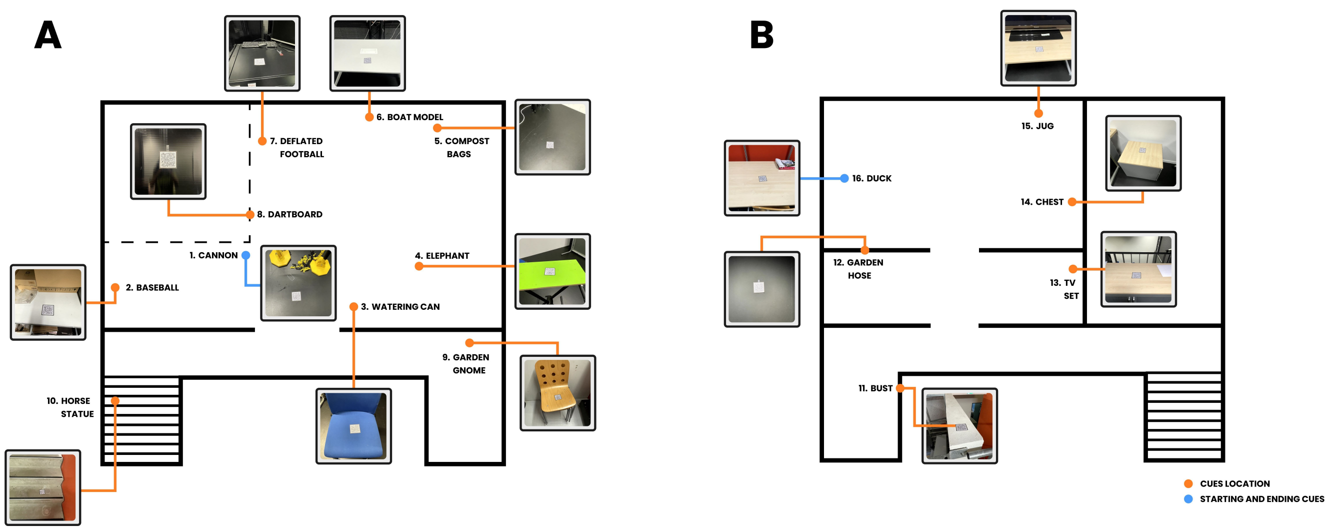


Figure 15: Experiment 2: overview of the physical environment where the QR codes were placed, and the path took by the participants. A and B are two different floors.

A.2 Stimuli Set and Selection Procedure

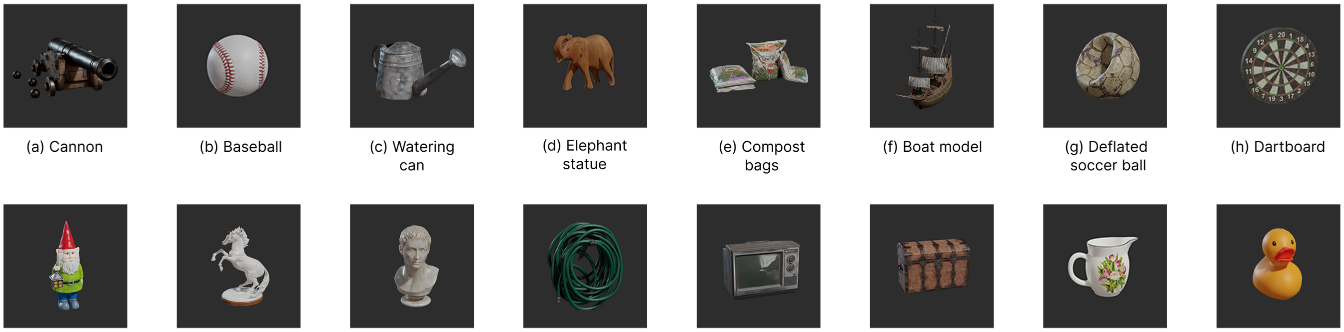
The stimuli we used in our two experiments are shown in Figure 16. We detail the procedure we used to select them.

We did not find any standard set of 3D objects for memory experiments, so we came up with our own set. During the early stages of our experiment design, we decided to use 16 stimuli, which is enough to allow their inclusion as a random effect in our statistical models, while keeping our experiment duration and complexity within reasonable limits. We first manually preselected a set of 30 objects from polyhaven.com, a public library of 3D models. We picked objects from three categories (tools, electronics, decorative), and chose objects that are neither too large (e.g., no furniture) nor too small.

In order to narrow down our selection and come up with a set of stimuli that are not too heterogeneous, we ran two pretests. In the first pretest, we recruited 30 French-speaking participants[10] via a crowdsourcing platform and asked them to rate each 3D model according to their valence (i.e., whether they elicited negative, neutral, or positive emotions) and arousal (i.e., how intense the emotion is). The reason was to discard objects that are too extreme in the emotions they elicit. We also asked participants to name the objects, and rate how easy they were to name. The reason was to discard objects that are difficult to name, because we will need participants to name them when they report their thoughts.

In the second pretest, we recruited 19 French speakers among our personal contacts, and asked them to rate the degree to which they

[10] As our study was conducted in France, stimuli were named in French.



(a) Cannon (b) Baseball (c) Watering can (d) Elephant statue (e) Compost bags (f) Boat model (g) Deflated soccer ball (h) Dartboard (i) Garden gnome (j) Horse statue (k) Bust (l) Garden hose (m) TV set (n) Chest (o) Jug (p) Plastic duck

Figure 16: The 16 stimuli used in experiment 1 and experiment 2.

would expect to find each object in each of our two experimental environments (see Figure 14 and Figure 15). For example, we asked participants if they would be surprised or not to find a baseball if they walked through the building corridor. We showed a photo and a description of each environment. All participants but three were already familiar with the first and second environment. The reason for this second pretest was that we wanted the 3D objects to be rather unexpected in order to make them memorable, but we also wanted to avoid outliers, e.g., objects that are extremely incongruous or totally expected.

For each of the 30 candidate objects and each of the 5 measurements (valence, arousal, nameability, expectedness in environment 1, expectedness in environment 2), we computed the mean response and the z-score. The z-score captures how much the object's mean deviates compared to other objects. All candidate objects whose mean was more than 2 standard deviations of the grand mean on any of the measurements (z-score less than -2 or more than 2) were excluded, with the exception of nameability where only negative z-scores were excluded. Examples of excluded objects were a guitar whose valence was too positive compared to the remaining objects, a megaphone whose arousal was too high, a trowel whose nameability was too low, and a house plant whose expectedness in environment 1 was too high. We were left with 20 stimuli, from which we manually selected 16 based on their size, in order to further filter out large and small objects. The final set of stimuli is shown in Figure 16. A table with all the pretest data can be found in the OSF repository.

A.3 Questionnaires

After the audio task in experiment 1, participants were asked to fill the following short questionnaires:

A *source identification questionnaire* adapted from Hoffman et al. [37], where the participant was shown 32 words, half of which were words describing one of the 16 stimuli (see Figure 16), and the other half of which were new words. They were asked whether each word referred to a stimulus they previously saw, and if so, if it was displayed in AR or not. For each word, they rated their confidence in their answer on a 7-point scale.

A *mind-wandering spontaneous scale*, translated in French from Driebergen [28], a self-report scale assessing individual differences in spontaneous mind-wandering. This state is characterised by the spontaneous generation of thoughts when the individual is 'zoning out'.

A *general intellectual humility scale*, translated in French from Leary et al. [47]. A scale for self-assessment of one's own intellectual humility, used to recognise and accept the limits of one's knowledge, understanding or beliefs.

A *computer experience questionnaire*, adapted and translated in French from Moffat et al. [58], our adapted version assess familiarity with smartphone and XR techonologies.

A simple *demographic questionnaire* collecting data on gender, age, level of education, and familiarity with the environment for experiment 1.

The purpose of these five questionnaires was to characterize the study sample and provide the basis for possible exploratory analyzes. They were given at the end in order to avoid possible priming and stereotypical threat effects [64].

An additional *deception check questionnaire* asked participants if they had doubts about the purpose of the study, and if so, to describe their doubts, in order to check that participants did believe in the cover story [17]. They were also asked if they reported a stimulus in the audio task and if they remembered it voluntarily. If they did, we excluded the stimulus.

Finally, the experimenter explained the true purpose of the study, and orally asked the participant if they had involuntary memories of experimental stimuli that they did not report during the audio task. As we further discuss in the results section, some participants reported that they did but thought they were not supposed to report them (recall the task instructions were purposefully vague and simply asked participants to report thoughts).

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009