Proactive and retroactive effects of novelty and rest on memory

Schomaker, J.; Raza, S.; Quent, J.A.; Anderson, M.

PROACTIVE AND RETROACTIVE EFFECTS OF NOVELTY AND REST ON MEMORY.

Sumaiyah Raza[1]
Judith Schomaker[3,4]
Jörn Alexander Quent[5,6,7]
Michael C. Anderson[1]
Richard N. Henson[1,2]

[1]*MRC Cognition & Brain Science Unit, University of Cambridge, Cambridge, UK.*
[2]*Department of Psychiatry, University of Cambridge, Cambridge, UK.*
[3]*Department Health, Medical and Neuropsychology, Leiden University, Leiden, The Netherlands.*
[4]*Department for Social and Behavioural Sciences, Leiden University, Leiden, The Netherlands.*
[5]*Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai, China.*
[6]*Key Laboratory of Computational Neuroscience and Brain-Inspired Intelligence (Fudan University), Ministry of Education, China.*
[7]*MOE Frontiers Center for Brain Science, Fudan University, Shanghai, China.*

Proactive and Retroactive Effects of Novelty and Rest on Memory.

Abstract

Novel experiences appear to benefit memory for unrelated information encoded shortly before or after the novel experience, in both rodents and humans. In contrast, other research has suggested that memory is impaired when encoding is followed by an effortful task, as opposed to simply resting. This apparent discrepancy in the literature may explain why a recent registered report by Quent and Henson (2022) found evidence against a retroactive novelty-related memory enhancement effect in humans, as the detrimental influence of the effortful nature of novel exploration may counteract the beneficial effect of its novelty. The present study therefore aims to explicitly test proactive and retroactive effects of novel exploration, and retroactive effects of wakeful rest. We will adapt a virtual-reality paradigm developed by Schomaker et al. (2022), to compare four groups of participants created by crossing novel versus familiar exploration with the study of words shortly before (Retroactive groups) or after (Proactive groups) that exploration. A fifth "wakeful rest" group will perform an easy auditory/visual detection task both before and after studying words. Memory will be tested with immediate free recall, delayed free recall, and delayed recognition. We will use Bayes Factors to assess evidence for the hypotheses that: 1) novel relative to familiar exploration will proactively benefit memory for the words, and 2) that wakeful rest will retroactively benefit memory relative to familiar exploration. Both novel exploration using virtual-reality and wakeful rest represent potential interventions for improving memory, with implications for clinical and ageing populations.

Introduction

Research with rodents has suggested that novel experiences can facilitate long-term memory for unrelated stimuli that are encoded close in time to the novel experience. Typically, rats are exposed to a weak version of a training protocol, such as contextual fear-conditioning, inhibitory avoidance training, or spatial object recognition (e.g. Moncada & Viola, 2007; Ballarini et al., 2009; Justel et al., 2021). If the rats also explore a novel environment shortly before or after the training, they show improved memory for this training when tested 24 hours later, compared with rats who did not explore a novel environment. This novelty-related memory enhancement has also been called 'behavioural tagging' (Moncada & Viola, 2007; Moncada et al., 2015). According to

Behavioural Tagging Theory (BTT) (Moncada & Viola, 2007; Moncada et al., 2015; see Dunsmoor et al., 2022 for an in-depth review), novelty induces plasticity-related neurochemical changes which facilitate consolidation of unrelated encoding that happens to occur in close temporal proximity to the novel experience.

Novel experiences also seem to have a comparable effect in humans. For example, Schomaker and colleagues developed a paradigm in which participants explore a novel virtual-reality (VR) environment shortly before encoding some words. They found that recall of these words after a short delay was better when participants had explored a novel environment compared to an environment with which they had previously been familiarised (Schomaker et al., 2014; Schomaker & Wittman, 2021; Schomaker et al., 2022). The memory enhancement was not accounted for by differences in subjective arousal following exploration. In another, more ecological example (Ballarini et al., 2013), school children were tested on their memory for a story heard the previous day. Children who had received a novel lesson 1 hour before or 1 hour after hearing the story were found to have better memory for it than children who did not receive a novel lesson. The memory benefit was not seen when the children had been familiarised with the lesson twice before. Nor was the benefit seen in children who received the novel lesson 4 hours before or after hearing the story, consistent with the plasticity-related changes induced by the novel experience only lasting for a limited time window. The novel lessons were conceptually unrelated to the story content, and the effect was comparable whether the novel lesson was about science or music, suggesting that the content of novel experiences is not important. With a similar paradigm, Butavand et al. (2020) showed that the memory improvement associated with novel lessons persisted for up to 45 days, supporting the robustness of the effect and the potential utility of novel experiences as a tool for improving memory. Taken together, these results support the existence of novelty-related memory enhancement in humans, in keeping with rodent research.

However, using a paradigm similar to that of Schomaker and colleagues, a recent registered report by Quent and Henson (2022) failed to replicate the effect of a novel VR experience on memory for words. During an initial, incidental study phase, participants made semantic (deep) or orthographic (shallow) judgments about a number of words. This was followed by a novel experience with immersive VR (iVR), in which participants used a headset to look around a virtual kitchen to find 20 objects, and their memory for the location of these objects was subsequently

tested (none of the participants had used iVR before). Immediately after the iVR, participants freely recalled the words studied before the iVR experience, and then the next day, completed a recognition test for the words, combined with a “remember/know” judgment to separate recollection and familiarity-based memory (Tulving, 1985). A separate group of participants completed the same procedure, except they had performed the identical iVR task on the preceding day, so that it was less novel. Bayes Factors showed evidence for no difference between groups in memory for the words, in either the immediate recall or delayed recognition test (collapsed across deep/shallow study task and across recollection or familiarity, with no evidence for or against any interactions with these factors). This suggests that, at least in this version of the paradigm, the novel iVR experience did not retroactively enhance memory, challenging previous research.

However, there are some key factors which may explain the divergent results of Quent and Henson (2022) and Schomaker et al. (2014). For example, a later study by Schomaker et al. (2021) suggested that active navigation within virtual environments (VEs) is important for the novelty effect (consistent with the original animal studies, which allowed rodents to explore the novel environment). Furthermore, Schomaker et al. (2022) found that exploration behaviour within their relatively large VEs, indexed by roaming entropy, predicted recall performance. By contrast, the single, small (~5vm x 4vm) room used by Quent and Henson (2022) required negligible navigation/ exploration. However, this finding may be specific to novel ‘spatial’ experiences, as a number of other studies have found a benefit of novelty with no requirement for exploration (e.g. Ballarini et al., 2013; Bunzeck & Duzel, 2006; Fenker et al., 2008; Abrahan et al., 2020). Nevertheless, in the experiment planned here, we intend to use a ‘spatial’ novel experience, and so we will employ the relatively large VEs used by Schomaker et al. (2022).

A second factor offered by Quent and Henson (2022) to explain the divergent results – and the one most relevant to the current proposal – is that, whereas the Schomaker et al. studies tested (and found) a proactive effect of novelty, Quent and Henson tested (and failed to find) a retroactive effect. Similarly, two other studies failed to find a retroactive effect of novelty in adults (McClay & Dunsmoor, 2018) and typically developing children (Baumann et al., 2020; though note they did find a retroactive novelty effect in those children with attention deficit hyperactivity disorder). Despite the previous findings in rodents and school children of both pro- and retro-active effects, it is possible that novelty only exerts a proactive effect in this type of VR paradigm.

Indeed, one reason for the lack of a retroactive effect of novelty may be an additional, counter-acting, detrimental effect of performing a cognitively demanding task shortly after encoding the words. This is because another literature has suggested that cognitively demanding tasks impair consolidation of previously encoded memories, relative to wakeful rest, as briefly reviewed next.

Several previous studies have shown that effortful tasks following encoding can impair memory. For example, Dewar et al. (2012) found that older adults had worse story recall when they engaged in an effortful visual discrimination task for 10 minutes following encoding, compared with when they rested for 10 minutes. This performance difference was evident after delays of 15 minutes, 30 minutes and 7 days. Similar beneficial effects following 10 minutes of post-encoding rest have also been reported in young adults (Craig et al., 2014), patients with amnesia (Cowan et al., 2005; Dewar et al., 2009) and patients with amnestic mild cognitive impairment (Alber et al., 2014). It seems unlikely that this benefit was due to intentional rehearsal of encoded information during the resting period, as the effect also occurred for recognition of 'unrecallable' words (i.e. foreign names; Dewar et al., 2014). Furthermore, the fact that the effortful tasks used in these studies were non-verbal and unrelated to the encoded information suggests that impaired memory reflects the cognitive load of the task, rather than interference from competing stimuli (e.g. at retrieval). A meta-analysis of ten similar studies found a significant, moderately-sized benefit of post-encoding rest on verbal memory (Cohen's d=0.38). These retroactive effects of cognitive effort are often explained in terms of consolidation theory: the theory that memory traces gradually become transformed after encoding, through cellular structural changes and system reorganisation, and until they are consolidated in this way, they remain susceptible to disruption (Miller & Pilzecker, 1900; Squire et al., 2015). Assuming these consolidation processes require cognitive/neural resources, then they will be impaired by effortful tasks (Wixted, 2004). As the novel VR task used by Quent and Henson (2022) was quite effortful (involving intentional learning of the location of objects in the room), it may have impaired consolidation of memories for the preceding words, counteracting any benefit of the novelty of the iVR task. Even if active navigation around Schomaker et al.'s larger VE is also effortful, because consolidation processes are 'asymmetrical' (can only occur after encoding), no such masking of the novelty effect would occur in Schomaker et al.'s proactive paradigm.

To test this possibility, we will compare the effects on memory of i) exploring a novel VE, ii) exploring a VE that has been familiarised the previous day and iii) spending the same amount of

time in “wakeful rest”, i.e. performing a very undemanding, unrelated task. Furthermore, we will explicitly test both proactive and retroactive effects of novelty: some participants will explore the VE before studying a list of words (“proactive” groups, as in Schomaker et al.’s studies), whereas others will explore the VE after studying a list of words (“retroactive” groups, as in Quent & Henson’s study). To match retention interval before testing memory, the proactive groups will perform the wakeful rest task after study, while the retroactive groups will perform the wakeful rest task before study (see ahead to Figure 1). Thus, there will be five groups in total: proactive novelty group, retroactive novelty group, proactive familiarised group, retroactive familiarised group, and the “wakeful rest” group who will perform the undemanding task both before and after studying words.

Memory will be tested with immediate free recall, 24-hour delayed free recall, and 24-hour delayed recognition with confidence ratings. The admission of both immediate and delayed memory tests will allow us to address a key difference between rodent and human research on novelty effects. Rodent studies typically test memory after an extended delay (e.g. 24 hours), based on the BTT assumption that protein-synthesis dependent consolidation processes that are facilitated by novelty, take place over several hours after encoding (Moncada et al., 2015). However, many human studies (e.g. those by Schomaker et al.) test memory after only short delays (e.g. 10 minutes). If novel exploration or wakeful rest affect consolidation processes, it is possible that their effects on memory will be greater on delayed tests. However, in case free recall after 24 hours is too close to floor, we will also test delayed recognition. Furthermore, we will use confidence ratings on the recognition memory task to estimate recollection versus familiarity processes, by applying the independent, dual-process model of Yonelinas (1994) to the resulting Receiver Operating Characteristic (ROC) curves. This distinction of recollection versus familiarity might be important because BTT claims that the novelty-induced neurochemical changes must occur in the same neural population that is encoding the stimuli (Moncada & Viola, 2007; Moncada et al., 2015). Given that exploration of novel environments is likely to engage the hippocampus, and that the hippocampus is associated with recollection but not familiarity (Aggleton & Brown, 1999; Argyropoulos et al., 2022), the effects of novelty could be found on estimates of recollection but not of familiarity. This is consistent with previous studies (e.g. Schomaker et al., 2014) that have found a novelty effect on recall but not recognition, on the assumption that their recognition data was dominated by familiarity.

In summary, we will employ a between-participant, two-way factorial design that crosses novel versus familiar exploration with pro- versus retro-active order, plus an additional control group experiencing wakeful rest only. Our dependent variables will be 1) immediate free recall, 2) delayed free recall and 3) delayed recognition memory split by recollection versus familiarity. Our primary aims are to: 1) replicate Schomaker et al.'s (2014) findings that exploration of a novel versus familiar VE will benefit immediate recall of unrelated words; and 2) replicate Dewar et al.'s (2012) findings that wakeful rest will retroactively benefit immediate and delayed memory compared to the more effortful task of exploring a familiar VE.

Regarding our first aim, BTT predicts a novelty-related memory benefit in both proactive and retroactive groups, and this benefit should be greater on memory tests that are delayed and that involve recall or recollection. Regarding our second aim, consolidation theory predicts that wakeful rest will benefit memory retroactively, but not proactively, and this effect should again be greater for delayed free recall or recollection during a delayed recognition test, assuming only hippocampally-dependent memory (rather than familiarity-based memory) undergoes consolidation (McClelland et al., 1995). Importantly, if both theories are true, the retroactive effect of novelty might be small or absent, owing to the difference between novel and familiar VEs being masked by the detrimental effect of performing any effortful task after encoding (and potentially explaining the lack of a retroactive novelty effect in Quent & Henson, 2022).

Finally, after the immediate recall test, for the groups experiencing the exploration interventions we will assess participants' subjective feelings of immersion/presence within the VEs. This may address another apparent discrepancy in the literature: Schomaker et al. (2014) found that these ratings positively predicted recall performance in their proactive paradigm, whereas Baumann et al. (2020) found these ratings negatively predicted recall in their retroactive paradigm. One possibility is that, in the proactive paradigm, greater immersion in the VE led to greater improvements in arousal and motivation during the subsequent encoding and test phases, in turn leading to better memory. Conversely, in the retroactive paradigm, greater immersion (which has been suggested to involve greater recruitment of cognitive resources; Barreda-Ángeles et al., 2021) may have caused greater impairment of consolidation after encoding and thus poorer memory performance, which would be in line with our theorised interaction between novelty and cognitive effort. We will also measure reaction times (RTs) during the encoding tasks (pleasantness

judgements), in case these are speeded by increased arousal following novel exploration (in the proactive group).

In summary, the results of this study will test the claims of neurocognitive theories like BTT and consolidation theory, thereby improving our understanding of the mechanisms underlying long-term memory, as well as evaluate two potential interventions for improving memory; important for potential clinical and educational application.

Procedure

Informed consent will be obtained, and participants will be tested online, once on each of 3 successive days. Participants will be randomly assigned to one of the five experimental groups. They will be instructed to complete all sessions of the experiment in the same quiet setting, at the same time each day. At the start of the experiment, they will report the type of setting they are in by selecting from some options (i.e. familiar, quiet / novel, unfamiliar / familiar, distractions present). The experiment will only continue if they are in a familiar setting, free from distraction, to reduce extraneous novelty effects. The order and duration of experimental tasks for each group is depicted in Figure 1. Detailed instructions for each task will be given at the start of each session. These initial instructions will include some comprehension tests. If participants fail a comprehension check twice, they will be instructed to contact the experimenter to help them understand the instructions before taking part. To discourage participants from multi-tasking, they will be prompted to complete the experiment in full-screen mode, and an alert box will appear should they exit this mode.

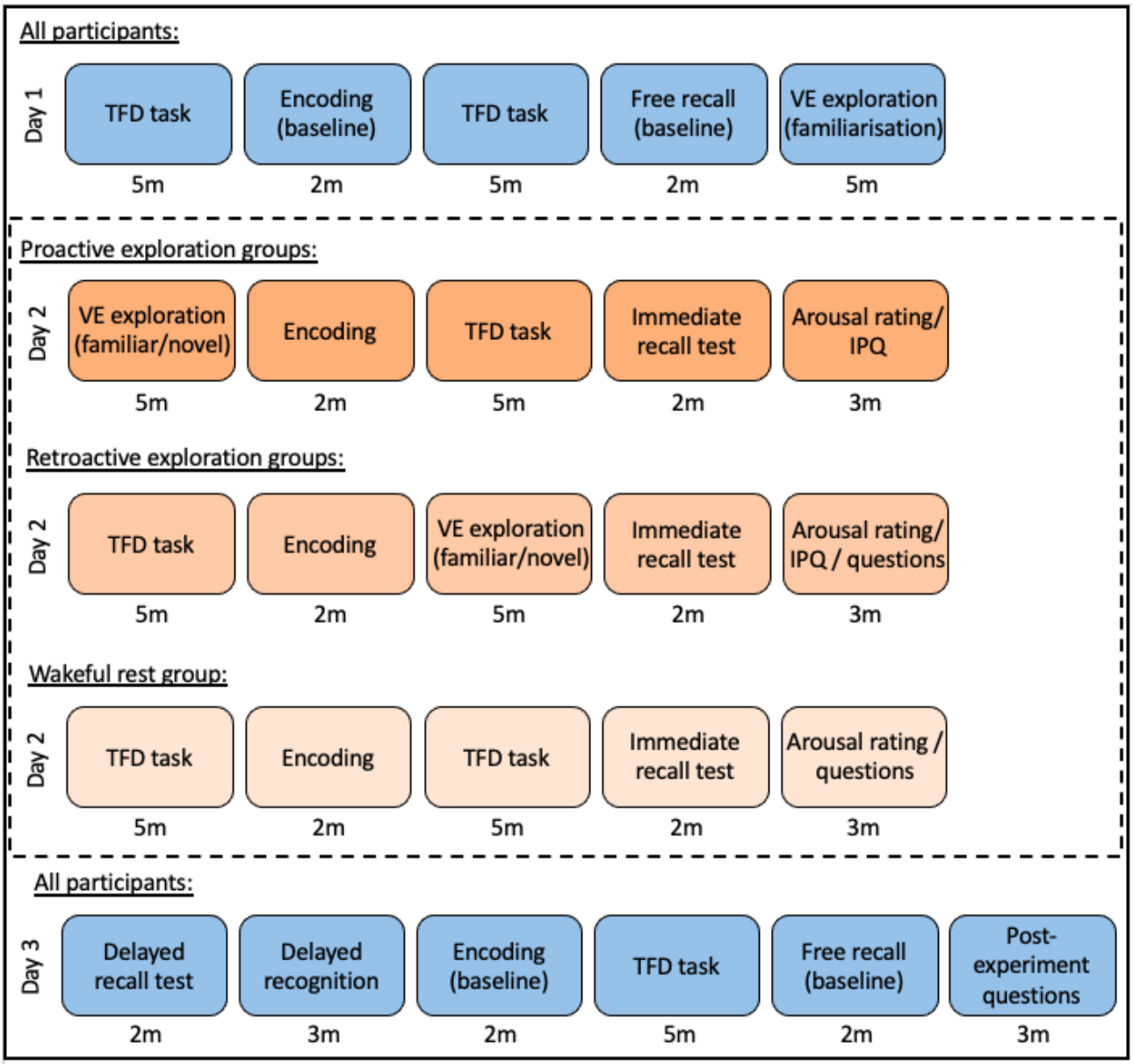
All participants:
Day 1
TFD task
5m
Encoding (baseline)
2m
TFD task
5m
Free recall (baseline)
2m
VE exploration (familiarisation)
5m
Proactive exploration groups:
Day 2
VE exploration (familiar/novel)
5m
Encoding
2m
TFD task
5m
Immediate recall test
2m
Arousal rating/ IPQ
3m
Retroactive exploration groups:
Day 2
TFD task
5m
Encoding
2m
VE exploration (familiar/novel)
5m
Immediate recall test
2m
Arousal rating/ IPQ / questions
3m
Wakeful rest group:
Day 2
TFD task
5m
Encoding
2m
TFD task
5m
Immediate recall test
2m
Arousal rating / questions
3m
All participants:
Day 3
Delayed recall test
2m
Delayed recognition
3m
Encoding (baseline)
2m
TFD task
5m
Free recall (baseline)
2m
Post-experiment questions
3m

Figure 1. Order of experimental tasks for a) the proactive groups and b) the retroactive groups. Day 1 will consist of the first baseline encoding and free recall phase, followed by familiarisation of a VE. Day 2 will involve one of the interventions (exploration of novel VE, exploration of familiar VE or wakeful rest/tone and flash detection (TFD) task, depending upon the participant's experimental group), encoding, and free recall test phases. Encoding will occur prior to VE exploration for the retroactive group and after VE exploration for the proactive group. This session will end with a subjective rating of arousal, questions about how often participants thought about the words during the intervention, and the IPQ for the exploration interventions only. Day 3 will begin with a delayed free recall test for the words learned on the previous day, followed by a recognition test for these words, a second baseline encoding and free recall phase, and finally some questions regarding participants' lifestyle and subjective experience of the experiment. Delay periods throughout are filled with the TFD task to match the time between encoding and test across proactive/retroactive groups, and the task structure across days. There will be approximately 5 minutes of task instruction at the start of each session, and an opportunity to report technical issues at the end of each session.

During encoding phases, participants will be asked to view words sequentially and decide for each whether they find the word to be pleasant or unpleasant. They be instructed that there are no correct or incorrect answers. Words will be presented in a random order, for 3000ms each (300ms intertrial interval), and pleasant/unpleasant responses will be given using the 'F'/'J' keys. Throughout the task, words will be presented inside a coloured box – the colour will differ for each of the 3 encoding tasks participants will undertake. Participants will be informed that their memory for words will be tested, but that they should focus on the task shown on-screen, to minimise the likelihood that they may cheat by writing down words or employ their own strategies for memorising words. The task will begin with four practice trials. One word-list will be used for each encoding phase and the assignment of lists will be counterbalanced across participants for each encoding phase. Free recall test phases will see participants type in as many words from the previously learned list as possible, within 2 minutes. The same coloured box from the encoding phase of the previously learned list will be shown while they type in their responses, to help reduce proactive interference from earlier encoding phases. The number of correctly recalled words from the list associated with each test will be counted. Words reported with typing errors

(up to 2 letters incorrect or mixed up) will be coded as correct unless the error results in a different meaningful word (e.g. 'well' given in place of 'wall')[1].

The tone and flash detection (TFD) task, which represents the undemanding "wakeful rest" task and will be used during retention periods, is adapted from Dewar et al. (2010). Participants will be instructed to relax, listen to the 'waterfall' sound, gaze at a central fixation cross on the screen, and press the space bar when they hear piano notes or see the fixation cross temporarily change colour. This low-effort task will reduce the extent to which participants may rehearse words or engage in autobiographical thinking which may introduce additional interference (Craig et al., 2014). It is possible that this task will introduce small deleterious effects on memory compared to complete rest, but, given that participants will be tested unsupervised, this task affords greater control and a way of checking that participants are not engaged in other activities. Participants will also be reminded how important it is to focus on this task, rather than check their phones for example, in order for their data to be informative (we believe the majority of participants want their data to be useful). A similar tone-detection task has been shown not to significantly disrupt delayed recall performance compared with complete wakeful rest in healthy adults (Dewar et al., 2010).

During VE exploration (VE familiarisation on Day 1 and the novel and familiar interventions on Day 2 for four of the groups), participants will explore one VE for 5 minutes, moving forward using the W key and change heading direction using the mouse or trackpad. They will be instructed to try to stick to paths. Participants in all groups will complete VE familiarisation on Day 1 (half of the participants will be familiarised with one VE and half with the other VE). For participants undergoing the novel and familiar exploration interventions, one of the TFD tasks before or after encoding (proactive / retroactive groups) will be replaced with VE exploration. The familiar group will explore the familiarised VE for 5 minutes and the novel group will explore the unseen VE.

Given the 6 possible orders in which the 3 word-lists may be encoded, combined with the counterbalancing of VE familiarisation, groups require multiples of 12 participants.

[1] If we encounter words that we feel should be handled differently from this pre-specified criterion, we will conduct a reanalysis to check if the results differ depending on the scoring criteria used.

At the end of the session on Day 2 1, participants will answer some questions relating to the intervention they experienced (either VE exploration or the TFD task). All will rate their arousal following the intervention using a sliding scale from 1 to 9 (1 = very calm, 9 = very excited). Participants in the proactive intervention groups and in the wakeful rest group will rate how often they thought about the words they learned during the intervention/TFD task. Participants who experienced the exploration interventions will also answer the 13-item IPQ (Schubert et al., 2001) to measure their subjective feelings of immersion in the VR experience.

During the recognition test, participants will be shown the list of 16 words encoded on the previous day (Day 2), interspersed with 16 lures. Words will be shown sequentially, in a random order, and will remain on screen whilst participants indicate how confidently they do or do not recognise a word (i.e. confident new/unsure new/unsure old/confident old). For estimating recollection and familiarity from responses in the recognition task, we will use a well-established model of Yonelinas (1994), to fit a two-parameter model to the ROC curve, where one parameter reflects the probability of recollection and the second is a signal-detection measure of continuous familiarity.

To conclude the experiment, participants will answer some questions about their lifestyle / subjective experience of the experiment (e.g. 'How often do you play online first-person view games?'). This data may provide additional insight into potential mediators of the Novelty effect.

Stimuli

The exploration interventions will utilise two VEs from Schomaker et al. (2022) consisting of fantasy lands. These were created in Unity Version 2017.2.21f1 (Unity Technologies, 2017) and matched for size, path length, and number of intersections. Each environment contains marked paths through brightly coloured foliage and unusual landmarks (Figure 2). Participants will download the Unity files and run them from their own laptop or PC. During exploration, the 3D coordinates of the moving agent within the VE will be logged for all timepoints with a sampling rate of 15 Hz. The TFD task (adapted from Dewar et al., 2010) for the wakeful rest periods will use a 5-minute recording of 'brown noise' (sounds like a waterfall) with 3-6 piano notes embedded and 3-6 screen colour changes occurring at random time-points. For the word encoding tasks, 64 English concrete nouns taken from Otten et al. (2001) will be divided into 3 lists of 16 words, plus

16 lures for use in the final recognition test. The lists have been selected so as not to differ in terms of characteristics available in the MRC Psycholinguistic Database (Wilson, 1987).

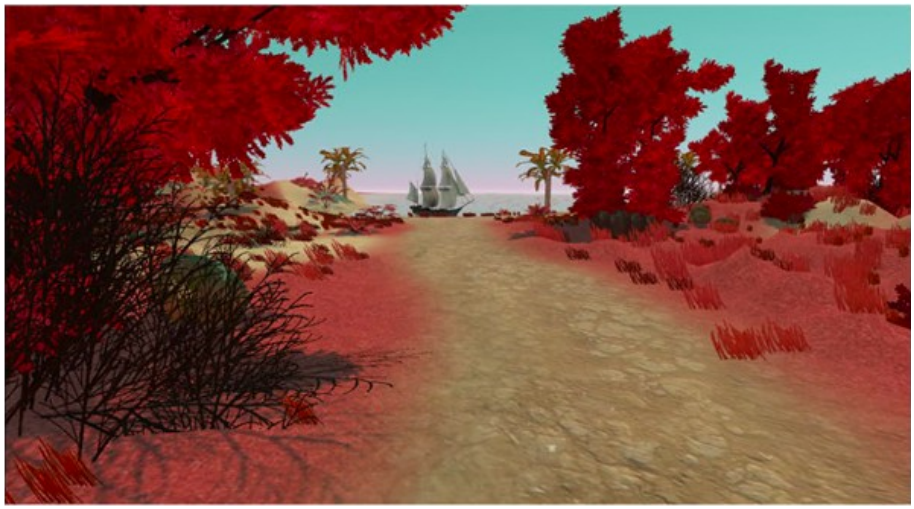

Figure 2. One of the VEs used in the exploration interventions (with thanks to Milan van der Kuil for VE design).

Statistical Design and Hypotheses

Inferences will be based on Bayes Factors (BFs), for the alternative (H1) versus null (H0) hypotheses, calculated for T-tests using the "ttestBF" function provided by the BayesFactor package in R with a 'medium' value for the "rscale" hyperparameter (default √2/2). We regard a conclusive outcome as a BF for H1 ("BF10") or BF for H0 ("BF01") that exceeds 6, based on the journal criteria. All registered hypotheses will be tested with directional, unpaired T-tests unless otherwise specified.

Our two main hypotheses are as follows: 1) 'Proactive Novelty effect' – that recall will be better in the proactive novelty group versus the proactive familiarised group (i.e. replicate Schomaker et al., 2014; and support BTT), and 2) 'Retroactive Resting effect' – that recall will be better in the wakeful rest group versus the retroactive familiarised group (i.e. replicate Dewar et al., 2012; and support Consolidation theory). These main hypotheses will be evaluated with immediate free recall performance (i.e. group mean number of correct recalls on the Day 2 test). We will terminate our sequential design (see Participants Section below) when conclusive evidence is found for both of these hypotheses (i.e., BF>6 for either the null or alternative). Once we terminate testing – either because of the above stopping rule or because we have recruited the maximum realistic sample size (see below).

After terminating data collection, we will conduct secondary analyses examining the 'Retroactive Novelty effect' (Hypothesis 3) by comparing immediate recall in the retroactive novelty group with the retroactive familiarised group. If a bidirectional Novelty effect exists, then recall should be better in the novelty group, however, if a Retroactive Resting effect exists then there may be no difference in recall between these groups. This latter pattern would support our theorised interaction between novelty and cognitive effort and explain the divergent results of Schomaker et al. and Quent & Henson.

For analyses involving immediate free recall, we will also try subtracting each participant's baseline recall performance, averaged across the first and last days, to see if it improves sensitivity by removing individual differences in memory ability.

The Proactive Novelty effect, Retroactive Resting effect, and Retroactive Novelty effect will also be assessed for other measures of memory. The same pattern of results is expected for delayed free recall performance (group mean number of correct recalls on the Day 3 free recall test) and for recollection on the delayed recognition test. However, for familiarity on the recognition test, there should be no effect of Novelty or Resting as explained above (see Introduction Section).

Finally, if we do find a Novelty effect (proactive and/or retroactive), we will conduct regression analyses across participants to relate the size of this effect to measures of 1) exploration (i.e. "roaming entropy"; Schomaker et al., 2022), 2) subjective ratings of arousal, and 3) IPQ scores. Furthermore, if we only find evidence of a Proactive Novelty effect, we will check, using a directional, unpaired T-test, whether response times on the encoding task differ following the Novel versus Familiar intervention; significantly faster responses following the Novel intervention would suggest that increased arousal/motivation could be responsible for the Proactive Novelty effect.

Participants

We will employ a Bayesian Sequential Design whereby we will continually recruit participants until we obtain conclusive evidence for H1 or H0 (i.e. BF10 or BF01 > 6) for both of our two main hypotheses, or until we reach a pre-determined maximum sample size based on resource

constraints. In our case, our maximum sample size is 168 per group (i.e. N=840 participants total). To allow full counterbalancing (see Procedure), we will recruit participants in batches of 12 per experimental group.

To determine the 'power' of our design, we performed 10,000 simulations of two directional, unpaired, one-tailed, Bayesian T-tests (code adapted from https://github.com/LevanBokeria/cbu_bayesian_sequential_designs/tree/multiple_stopping_rules), for each of our hypotheses with effect sizes (1) taken from previous literature for H1, or (2) set to zero for H0. For Hypothesis 1 (Proactive Novelty effect), we set Cohen's d=0.44, based on half that of Schomaker et al. (2014). We halved this effect size to account for potential effect size inflation, e.g. due to publication bias, because the estimate is based on only one study. For Hypothesis 2 (Retroactive Resting effect), we set Cohen's d=0.38, based on a meta-analysis of 10 studies investigating post-encoding rest.

Our simulations (see Table 1) demonstrate that, with a maximum N of 168 per experimental group, our study is well-powered (i.e. ~80% power) to detect both of our main effects of interest if they both exist (Novelty effect/Resting effect). If neither effect exists, our study has a moderate chance of finding conclusive evidence for the null hypothesis in both cases (i.e. ~60% power). Our study has a ~85-95% probability of finding conclusive evidence for H1 or H0 for at least one effect (i.e. the sum of the probability of conclusive and partially conclusive evidence). Across all combinations of H1 and H0, our study has a low chance of producing misleading evidence (i.e. < 4% false positive rate). Furthermore, we are likely to reach conclusive evidence (i.e. terminate our experiment) before recruiting our maximum N of 840. For example, to obtain conclusive evidence of both H1s when both are true, the median number of participants required across all our simulations was 108 per group (total N=540). Similarly, to obtain conclusive evidence of both H0s when neither effect exists, the median number of participants required was 120 per group (total N=600).

Table 1. Bayesian power analysis. Percentage of simulations providing evidence of various types for our two main effects of interest (corresponding to the probability of obtaining evidence of various types in our study).

	Both effects exist. (d = 0.44 & 0.44)	Only the Novelty effect exists. (d = 0.44 & 0)	Only the Resting effect exists. (d = 0 & 0.44)	Neither effect exists. (d = 0 & 0)
Conclusive evidence for both hypotheses. [1]	79.71%	73.09%	65.05%	59.86%
Partially conclusive evidence. [2]	16.54%	21.89%	25.43%	26.53%
Inconclusive evidence for both hypotheses. [3]	1.28%	5.78%	2.39%	10.07%
Misleading evidence for either hypothesis. [4]	2.47%	2.63%	3.74%	3.54%

[1] Conclusive evidence refers to a BF exceeding criterion that is not misleading, e.g., in the case that both effects exist (column 1 of Table) this would equal BF10>6 for both effects; in the case that only the Novelty effect exists (column 2 of Table) this would equal BF10>6 for the Novelty effect and BF01>6 for the Resting effect).

[2] Partially conclusive evidence refers to a not-misleading BF exceeding criterion for one effect, and not exceeding criterion for the other effect, e.g., in that case that only a Resting effect exists (column 3 of Table), this could be a BF10>6 for the Resting effect and a BF not exceeding criterion for the Novelty effect or a BF01>6 for the Novelty effect and a BF not exceeding criterion for the Resting effect.

[3] Inconclusive evidence refers to a case where BFs for both effects do not exceed criterion.

[4] Misleading evidence refers to cases where one or both BFs exceeded criterion for the incorrect hypotheses compared with the effects actually in existence, e.g., where neither effect exists (column 4 of Table) this could be BF10>6 for one or both effects.

This study is of the type approved by the Cambridge Psychology Research Ethics committee (PRE.2020.018). Participants will be recruited using Prolific (www.prolific.co) and the MRC Cognition and Brain Sciences' SONA system, in-house participant panel. They will be paid at the end of the three days (i.e., end of Day 3) at a standard rate of £6/h with each session rounded up to the nearest 15 minutes (i.e., £3 per day = £10 total), plus a bonus of £5 after Day 3 to encourage completion of all sessions. All participants will need to self-report that they are 18-40 years old, have normal or corrected-to-normal vision, have normal hearing, are fluent in English, and have no history of diagnosed neurological or psychiatric illness.

Data Quality Checks

There are some circumstances under which a participant's involvement in the study will be terminated prior to completion. Ideally, participants will be tested over 3 consecutive days with 24 hours between each session, however, participants with a delay of up to 48 hours between the first and second sessions will still be included in the study. Those with a delay of longer than 30 hours between the second and third session, or not completing the study within 4 days will have their involvement terminated. Furthermore, at the end of each session participants will complete an attention check (as described by Oppenheimer et al., 2009) in which they must follow explicit instructions about how to respond to a simple question (i.e. 'When asked to choose a colour, you must select green. This is an attention check. Based on the text above, which colour do you choose? [Select from options green/blue/red/yellow]'). Participants' involvement in the study will be terminated if they fail 2 attention checks throughout the study. Termination will also occur if: accuracy is less than 70% on a TFD task, or if participants respond in less than 150ms to more than 70% of trials on an encoding task.

Individual participant's data will be excluded and replaced with a new participant if they spend less than 0.5s on instruction pages, take breaks exceeding 5 minutes during a single session, or if any of the following criteria fall 1.5x below the inter-quartile range (IQR) across all participants: total distance travelled during VE exploration, performance on the TFD tasks pooled across all of these tasks, overall recall performance pooled across the free recall tests on all days (exclusion will also be applied if the latter criterion is 1.5x above IQR as this may indicate participants 'cheating'). Data on issues experienced by participants, which participants will have the opportunity to report at the end of each session, will also be used to aid screening (i.e. participants may need to be excluded if they experienced major technical issues). Finally, at the end of the entire experiment, we will include a 'seriousness' check (e.g. Aust et al., 2013). This will remind participants that we depend upon having good quality data, prompt them to indicate how seriously they took part in the experiment and provide an opportunity for them to input any potential reasons that their data should not be used. Any participants with responses that suggest they were not properly attending to the experiment, or that they cheated in some way, will be excluded.

Acknowledgements

We thank Milan van der Kuil for the design of the virtual environments. This work and Sumaiyah Raza's PhD studentship (SUAG/096 G116768) is funded by the United Kingdom's Medical Research Council (SUAG/086 G116768; SUAG/079 G116768). It is also supported by the Dutch Research Council (406.XS.01.144), the China Postdoctoral Science Foundation (2022M720818) and the National Natural Science Foundation of China (32250410282).

The authors declare no conflicting interests. For open access, the authors have applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission.

References

Abrahan, V. D., Psyrdellis, M., & Justel, N. (2020). Novelty exposure modulates visual and verbal emotional memory: an experimental design with adults. *Acta Psychologica*, *205*, 103029.

Aggleton, J. P., & Brown, M. W. (1999). Episodic memory, amnesia, and the hippocampal–anterior thalamic axis. *Behavioral and brain sciences*, *22*(3), 425-444.

Alber, J., Della Sala, S., & Dewar, M. (2014). Minimizing interference with early consolidation boosts 7-day retention in amnesic patients. *Neuropsychology*, *28*(5), 667.

Argyropoulos, G. P., Dell'Acqua, C., Butler, E., Loane, C., Roca-Fernandez, A., Almozel, A., ... & Butler, C. R. (2022). Functional specialization of the medial temporal lobes in human recognition memory: Dissociating effects of hippocampal versus parahippocampal damage. *Cerebral Cortex*, *32*(8), 1637-1652.

Aust, F., Diedenhofen, B., Ullrich, S., & Musch, J. (2013). Seriousness checks are useful to improve data validity in online research. *Behavior research methods*, *45*, 527-535.

Ballarini, F., Martínez, M. C., Díaz Perez, M., Moncada, D., & Viola, H. (2013). Memory in elementary school children is improved by an unrelated novel experience. *PloS one*, *8*(6), e66875.

Ballarini, F., Moncada, D., Martinez, M. C., Alen, N., & Viola, H. (2009). Behavioral tagging is a general mechanism of long-term memory formation. *Proceedings of the National Academy of Sciences*, *106*(34), 14599-14604.

Barreda-Ángeles, M., Aleix-Guillaume, S., & Pereda-Baños, A. (2021). Virtual reality storytelling as a double-edged sword: Immersive presentation of nonfiction 360-video is associated

with impaired cognitive information processing. *Communication monographs*, *88*(2), 154-173.

Baumann, V., Birnbaum, T., Breitling-Ziegler, C., Tegelbeckers, J., Dambacher, J., Edelmann, E., ... & Krauel, K. (2020). Exploration of a novel virtual environment improves memory consolidation in ADHD. *Scientific Reports*, *10*(1), 1-15.

Biel, D., & Bunzeck, N. (2019). Novelty before or after word learning does not affect subsequent memory performance. *Frontiers in Psychology*, *10*, 1379.

Bunzeck, N., & Düzel, E. (2006). Absolute coding of stimulus novelty in the human substantia nigra/VTA. *Neuron*, *51*(3), 369-379.

Butavand, D.R., Hirsch, I., Tomaiuolo, M., Moncada, D., Viola, H., & Ballarini, F. (2020). Novelty improves the formation and persistence of memory in a naturalistic school scenario. *Frontiers in psychology*, *11*, 48.

Cowan, N., Beschin, N., Perini, M., & Della Sala, S. (2005). Just lying there, remembering: Improving recall of prose in amnesic patients with mild cognitive impairment by minimising interference. *Memory*, *13*(3-4), 435-440.

Craig, M., Della Sala, S., & Dewar, M. (2014). Autobiographical thinking interferes with episodic memory consolidation. *PloS one*, *9*(4), e93915.

Dewar, M., Alber, J., Butler, C., Cowan, N., & Della Sala, S. (2012). Brief wakeful resting boosts new memories over the long term. *Psychological science*, *23*(9), 955-960.

Dewar, M., Alber, J., Cowan, N., & Della Sala, S. (2014). Boosting long-term memory via wakeful rest: intentional rehearsal is not necessary, consolidation is sufficient. *PloS one*, *9*(10), e109542.

Dewar, M., Della Sala, S., Beschin, N., & Cowan, N. (2010). Profound retroactive interference in anterograde amnesia: what interferes?. *Neuropsychology*, *24*(3), 357.

Dewar, M., Garcia, Y. F., Cowan, N., & Sala, S. D. (2009). Delaying interference enhances memory consolidation in amnesic patients. *Neuropsychology*, *23*(5), 627.

Dunsmoor, J. E., Murty, V. P., Clewett, D., Phelps, E. A., & Davachi, L. (2022). Tag and capture: how salient experiences target and rescue nearby events in memory. *Trends in Cognitive Sciences*.

Fenker, D. B., Frey, J. U., Schuetze, H., Heipertz, D., Heinze, H. J., & Duzel, E. (2008). Novel scenes improve recollection and recall of words. *Journal of cognitive neuroscience*, *20*(7), 1250-1265.

Justel, N., Salguero, A., Marengo, L., Psyrdellis, M., & Pautassi, R. M. (2021). Open field exposure facilitates the expression of a spatial, recognition memory. *Neuroscience Letters*, *757*, 135997.

McClay, M., & Dunsmoor, J. E. (2018). Effects of post-training novelty exposure on contextual fear memory: An attempt to translate behavioral tagging to humans. *bioRxiv*, 479659.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychological review*, *102*(3), 419.

McGhee, J. D., Cowan, N., Beschin, N., Mosconi, C., & Della Sala, S. (2020). Wakeful rest benefits before and after encoding in anterograde amnesia. *Neuropsychology*, *34*(5), 524.

Merhav, M., & Rosenblum, K. (2008). Facilitation of taste memory acquisition by experiencing previous novel taste is protein-synthesis dependent. *Learning & Memory*, *15*(7), 501-507.

Moncada, D., & Viola, H. (2007). Induction of long-term memory by exposure to novelty requires protein synthesis: evidence for a behavioral tagging. *Journal of Neuroscience*, *27*(28), 7476-7481.

Moncada, D., Ballarini, F., & Viola, H. (2015). Behavioral tagging: a translation of the synaptic tagging and capture hypothesis. *Neural plasticity*.

Müller, G. E., & Pilzecker, A. (1900). *Experimentelle beiträge zur lehre vom gedächtniss* (Vol. 1). JA Barth.

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of experimental social psychology*, *45*(4), 867-872.

Otten, L. J., Henson, R. N., & Rugg, M. D. (2001). Depth of processing effects on neural correlates of memory encoding: relationship between findings from across-and within-task comparisons. *Brain*, *124*(2), 399-412.

Quent, J. A., & Henson, R. N. (2022). Novel immersive virtual reality experiences do not produce retroactive memory benefits for unrelated material. *Quarterly Journal of Experimental Psychology*, *75*(12), 2197-2210.

Schomaker, J., & Meeter, M. (2015). Short-and long-lasting consequences of novelty, deviance and surprise on brain and cognition. *Neuroscience & Biobehavioral Reviews*, *55*, 268-279.

Schomaker, J., & Wittmann, B. C. (2021). Effects of active exploration on novelty-related declarative memory enhancement. *Neurobiology of Learning and Memory*, *179*, 107403.

Schomaker, J., Baumann, V., & Ruitenberg, M. F. (2022). Effects of exploring a novel environment on memory across the lifespan. *Scientific Reports*, *12*(1), 16631.

Schomaker, J., van Bronkhorst, M. L., & Meeter, M. (2014). Exploring a novel environment improves motivation and promotes recall of words. *Frontiers in psychology*, *5*, 918.

Schubert, T., Friedmann, F., & Regenbrecht, H. (2001). The experience of presence: Factor analytic insights. *Presence: Teleoperators & Virtual Environments*, *10*(3), 266-281.

Squire, L. R., Genzel, L., Wixted, J. T., & Morris, R. G. (2015). Memory consolidation. *Cold Spring Harbor perspectives in biology*, *7*(8), a021766.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology/Psychologie canadienne*, *26*(1), 1.

Wilson, M. D. (1987). MRC Psycholinguistic Database: Machine Usable Dictionary: Version 2.00. Informatics Division, Science and Engineering Research Council, Rutherford Appleton Laboratory.

Wixted, J. T. (2004). The psychology and neuroscience of forgetting. *Annu. Rev. Psychol.*, *55*, 235-269.

Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: evidence for a dual-process model. *Journal of experimental psychology: Learning, memory, and cognition*, *20*(6), 1341.