

Alcohol-Induced Retrograde Facilitation?

Mixed Evidence in a Preregistered Replication and Encoding-Maintenance-Retrieval Analysis

J. Quevedo Pütter  and E. Erdfelder 

Department of Psychology, School of Social Sciences, University of Mannheim, Mannheim, Germany

Abstract: Somewhat counterintuitively, alcohol consumption following learning of new information has been shown to enhance performance on a delayed subsequent memory test. This phenomenon has become known as the retrograde facilitation effect (Parker et al., 1981). Although conceptually replicated repeatedly, serious methodological problems are associated with most previous demonstrations of retrograde facilitation. Moreover, two potential explanations have been proposed, the interference and the consolidation hypothesis. So far, empirical evidence for and against both hypotheses is inconclusive (Wixted, 2004). To scrutinize the existence of the effect, we conducted a preregistered replication that avoided common methodological pitfalls. In addition, we used Küpper-Tetzel and Erdfelder's (2012) multinomial processing tree (MPT) model to disentangle encoding, maintenance, and retrieval contributions to memory performance. With a total sample size of $N = 93$, we found no evidence for retrograde facilitation in overall cued or free recall of previously presented word pairs. In line with this, MPT analyses also showed no reliable difference in maintenance probabilities. However, MPT analyses revealed a robust alcohol advantage in retrieval. We conclude that alcohol-induced retrograde facilitation might exist and be driven by an underlying retrieval benefit. Future research is needed to investigate potential moderators and mediators of the effect explicitly.

Keywords: retrograde facilitation, alcohol, interference hypothesis, consolidation hypothesis, conceptual replication, multinomial processing tree (MPT) modeling



Alcohol is one of the most widely used and abused substances worldwide, with estimated prevalences of current drinking and heavy episodic drinking as high as 47% and 20%, respectively (Manthey et al., 2019). To better understand the consequences of drinking alcohol on everyday functioning, considerable effort has been devoted toward investigating cognitive alterations following acute alcohol intake (Mintzer, 2007). One particularly surprising observation originating from this research was that alcohol consumption following learning of new information apparently enhances performance on a delayed subsequent memory test. Parker et al. (1980) were the first to empirically demonstrate this so-called retrograde facilitation effect in a laboratory experiment using human subjects. While the effect seems to replicate across a variety of experimental paradigms, there are important methodological weaknesses associated with most of the studies used so far. Moreover, the question of which mechanisms underlie retrograde

facilitation has rarely been addressed empirically, and the corresponding research has not resulted in conclusive results so far (Wixted, 2004). Both issues are intertwined and of great importance for both theoretical and practical reasons. Therefore, in this registered report, we proposed a replication study that addresses them jointly. Specifically, we suggested a study design that avoids common methodological pitfalls by conceptually replicating the original procedure described by Parker et al. (1981) and, in addition, reanalyzing the replication data using a multinomial processing tree (MPT) model proposed by Küpper-Tetzel and Erdfelder (2012) to disentangle encoding, maintenance, and retrieval contributions to memory performance (for a review of MPT models, see Erdfelder et al., 2009; for an MPT tutorial, see Schmidt et al., 2023). Thus, we addressed two main goals in the present study: (a) rigorous conceptual replication of the retrograde facilitation effect to ensure that it is not an artifact of confounding variables and (b) using an established MPT measurement model to empirically assess the hypotheses proposed to explain retrograde facilitation. The preregistered study protocol is publicly available on the Open Science Framework (OSF; <https://doi.org/10.17605/OSF.IO/DK8FJ>; Quevedo Pütter et al., 2020).

Prior Research on Retrograde Facilitation

In 1981, Parker and colleagues demonstrated retrograde facilitation in a highly rigorous manner. In their experiment, alcohol versus placebo administration after learning was manipulated as a within-subjects factor. The retention interval lasted 7 hrs in each case, during which participants remained in a controlled laboratory environment, abstaining from any cognitively challenging tasks and further alcohol consumption. Participants showed significantly better performances in a recognition task after consuming either 0.5 or 1.0 ml alcohol per kg body weight (ml/kg) than after consuming 0.25 ml/kg or no alcohol at all. To illustrate, assuming a body weight of 80 kg, an alcohol weight of 0.8 g/ml, and an alcohol content of 5% for beer, the dose of 0.25 ml/kg would correspond to 20 ml (16 g) of pure alcohol or 0.4 l of beer, the dose of 0.5 ml/kg to 40 ml (32 g) of pure alcohol or 0.8 l of beer, and the dose of 1.0 ml/kg to 80 ml (64 g) of pure alcohol or 1.6 l of beer. Note that these quantities vary as a function of the body weights of individual participants and are thus examples only. Also note that beer is only used as a commonly known reference standard here. Parker et al. (1981) actually used a mixture of pure alcohol and a masking solution in their experiment.

In terms of internal validity, the Parker et al. (1981) study can be considered a convincing test of retrograde facilitation by alcohol. Most importantly, the retention interval of 7 h within a single day is (a) long enough to ensure that participants are completely sober when memory is tested and (b) short enough so that the interval does not cover a night of sleep. Unfortunately, other researchers have tried to avoid long retention intervals in controlled environments by either employing considerably shorter retention intervals (Knowles & Duka, 2004; R. E. Mann et al., 1984; Parker et al., 1980, study 1; Tyson & Schirmuly, 1994) or by dismissing participants after the learning phase and extending the retention interval up to 48 h (Bruce & Pihl, 1997; Carlyle et al., 2017; Lamberty et al., 1990; Mueller et al., 1983; Parker et al., 1980, study 2; Weafer et al., 2016a, 2016b). As detailed below, both approaches have considerable drawbacks compared to the original procedure used by Parker et al. (1981).

Short retention intervals involve the risk that participants are still intoxicated when memory is tested. To illustrate, R. E. Mann et al. (1984) conducted two separate studies with retention intervals of approximately 80 min and 140 min. They reported that participants were intoxicated to a considerable degree when memory was tested and rightfully concluded that memory tests conducted in a sober state for both the placebo and alcohol groups could have led to different results. One obvious criticism is that memory performance in the alcohol condition is hampered under such circumstances because

of state-dependent learning (Goodwin et al., 1969). Accordingly, memory performance of participants in the alcohol condition probably suffered from discrepant internal states during learning (sober state) and testing (intoxicated state). Nonetheless, statistically significant retrograde facilitation effects emerged in both studies of R. E. Mann et al. (1984). In principle, this finding is in line with a strong and robust retrograde facilitation effect. However, as participants in the alcohol condition were sober during learning and intoxicated during testing, the true size of the effect is probably underestimated, making it difficult to infer anything about its practical relevance. Even more importantly, as detailed below, one potential explanation of retrograde facilitation by alcohol – the interference hypothesis – can only be tested in a strict manner if accessibility of information in memory is not hampered by unrelated factors such as the state-dependency of memory.

An alternative approach to avoid long retention intervals in supervised controlled environments involves use of retention intervals extending to the next day. Hence, learning and memory testing take place on subsequent days, and participants can be sent home after the first session. This procedure has been employed by various researchers, with retention intervals ranging between 16 hrs (Carlyle et al., 2017) and 48 hrs (Weafer et al., 2016a, 2016b), always encompassing at least 1 night of sleep. While this approach avoids the problem of ongoing alcohol intoxication during recall, it comes with two other, possibly even more severe methodological challenges.

First, sleep research has shown that a presleep dose of alcohol typically increases slow-wave sleep (SWS) in the first half of the night for both healthy adults (Ebrahim et al., 2013) and adolescents (Chan et al., 2013). This poses serious problems for retrograde facilitation research because sleep, and especially SWS, has been argued to play a predominant role in the consolidation of declarative memories (Rasch & Born, 2013). Accordingly, it is possible that the apparent effects of retrograde facilitation by alcohol in these studies are actually due to memory consolidation during sleep. Alternatively, it could also be argued that effects of alcohol on sleep architecture may lead to a decline in memory performance. For example, an increase in sleep disruption following a presleep dose of alcohol (Ebrahim et al., 2013) could deteriorate participants' test performance on the following day due to sleep deprivation and increased daytime sleepiness during testing. However, this objection was taken into account in two studies by Weafer et al. (2016a, 2016b) who used a retention interval of 48 h to rule out that participants perform memory tests in a sleep-deprived or hungover state.

Second, acute alcohol intake has been shown to interact with sleep deprivation to increase daytime sleepiness

(Roehrs & Roth, 2001). Such an effect may be linked to the first problem, as daytime napping during the retention interval has been found to increase declarative memory performance (Tucker et al., 2006).

Mechanisms Proposed to Underlie Retrograde Facilitation by Alcohol

Two explanations for retrograde facilitation by alcohol have been proposed, the interference hypothesis and the consolidation hypothesis (Mueller et al., 1983; Tyson & Schirmuly, 1994). According to the interference hypothesis, retrograde facilitation is due to new incoming information being encoded and stored in memory less efficiently during intoxication, resulting in anterograde amnesia or at least mild forms thereof. As a consequence, memory representations originating from the previous learning phase are protected from retroactive interference. It has been shown convincingly that deterioration of memory following similarity-based retroactive interference is not due to distortion or loss of memory representations, but rather caused by impaired accessibility of information available in memory (e.g., Lohnas et al., 2015; Tulving & Psotka, 1971). By implication, the interference hypothesis predicts that retrograde facilitation results from enhanced accessibility of learned information in the testing phase, not from improved maintenance across the retention interval. In contrast, proponents of the consolidation hypothesis argue that alcohol actively boosts postlearning processing, resulting in more stable and durable memory representations (Parker et al., 1981), for example, by acting on brain regions involved in learning and memory such as the hippocampus (White, 1996) or by providing ideal (i.e., interference-free) conditions for undisrupted memory consolidation (Mednick et al., 2011; Wixted, 2004, 2010). Hence, retrograde facilitation emerges as a consequence of enhanced maintenance across the retention interval. In sum, while the interference hypothesis posits that alcohol-induced retrograde facilitation effects are due to better retrieval as a consequence of reduced retroactive interference, the consolidation hypothesis identifies better maintenance in memory as the main cause.

Both hypotheses have received some empirical support so far. Specifically, the interference hypothesis is in line with the findings by Mueller et al. (1983) and Tyson and Schirmuly (1994), respectively, showing that (a) retrograde facilitation is not time-dependent so that its strength does not depend on whether alcohol administration directly follows the learning phase or occurs later (although it has been shown that retroactive interference is also time-dependent to a certain degree; for a review, see Wixted, 2004) and (b) that retrograde facilitation is more

pronounced in memory tasks with few or no retrieval cues (e.g., free recall tasks) compared to tasks providing strong retrieval cues (e.g., recognition tasks), that is, effect sizes between alcohol and placebo conditions are more pronounced in free recall tasks than in recognition tasks. In contrast, Parker et al. (1981) favor a consolidation explanation because they observed a retrograde facilitation effect although a relatively low dose of alcohol was used in one condition (0.5 ml/kg) that is unlikely to decrease retroactive interference significantly. Perhaps more importantly, cognitive activity during the retention interval was reduced to a minimum in their study, making protection against interfering information obsolete.

Measurement of Memory Maintenance and Retrieval Contributions

Based on the evidence summarized in the previous section, Wixted (2004) concluded that attempts to empirically discriminate between the interference and the consolidation hypothesis have been inconclusive so far. He supposed that identification of the mediating physiological mechanisms might be a precondition for discriminating between the two theoretical accounts successfully. Following a different rationale, we propose an appropriate MPT model (see Erdfelder et al., 2009, for a review) to disentangle maintenance and retrieval contributions to the retrograde facilitation effect on a functional rather than a physiological level.

One particularly promising MPT model for our purposes is the encoding-maintenance-retrieval (EMR) model (Küpper-Tetzel & Erdfelder, 2012), an extension of the previously proposed storage-retrieval model by Rouder and Batchelder (1998). Küpper-Tetzel and Erdfelder's EMR model measures encoding, maintenance, and retrieval contributions to overall memory performance. It requires a study design that involves study of paired associates, immediately followed by a cued recall task. After a retention interval, additional free and final cued recall tests are administered (free-then-cued-recall paradigm). As detailed below, these methodological requirements are easily combined with the typical retrograde facilitation research design, enabling us to decide empirically whether retrograde facilitation by alcohol – if it exists – is driven by (a) improved maintenance across the retention interval, (b) improved retrieval in delayed subsequent free recall, or (c) both. Crucially, as shown above, the interference hypothesis and the consolidation hypothesis clearly differ with respect to which of these processes are expected to underlie retrograde facilitation. Also, because these hypotheses are not mutually exclusive, it is conceivable that both are correct, as reflected in improvements of both maintenance and retrieval after alcohol consumption.

Table 1. Twelve observable event categories E_1 to E_{12} for a study design employing an initial cued recall immediately following the learning phase and a free-then-cued-recall in the recall phase

Initial cued recall	Final cued recall	Final free recall		
		Both words	Exactly one word	Neither word
Correct	Correct	E_1	E_2	E_3
	Incorrect	E_4	E_5	E_6
Incorrect	Correct	E_7	E_8	E_9
	Incorrect	E_{10}	E_{11}	E_{12}

Note. Twelve possible memory test outcomes in the 3-level free-then-cued-recall paradigm proposed by Küpper-Tetzel and Erdfelder (2012).

Because the EMR model relies on a study design encompassing one initial cued recall test (2 possible outcomes per studied word pair: correct vs. incorrect), one later free recall test (3 possible outcomes: 0, 1, or 2 words of a pair recalled), and one final cued recall test (2 possible outcomes: correct vs. incorrect), a total of $2 \times 3 \times 2 = 12$ observable outcome patterns E_1 to E_{12} can occur for each word pair, depending on participants' performance in the three memory tests (cf. Table 1). The model comprises seven latent parameters: the probability of successful encoding of an association (e), the probabilities of maintenance of stored associations across the retention interval given successful versus unsuccessful initial cued recall (m_s and m_u , respectively), the probabilities of successful retrieval in free (r_f) and cued recall (r_c), and finally, the probabilities of single word retrieval in free recall given successful versus unsuccessful associative encoding (s and u , respectively). In total, the model includes 32 possible sequences of successful versus unsuccessful encoding, maintenance, and retrieval steps (so-called branches), each terminating in one of the 12 event categories E_1 to E_{12} . The probability of a branch is just the product of parameters along that branch, and the probability of a category equals the sum of the branch probabilities corresponding to this category. Based on these rules, model equations are obtained that represent the probabilities of the observable categories E_1 to E_{12} as functions of the seven model parameters. Given these equations and a set of event frequencies, model parameters can be estimated using either standard maximum likelihood techniques (Batchelder & Riefer, 1999; Hu & Batchelder, 1994; Moshagen, 2010) or Bayesian estimation techniques for hierarchical model versions that account for individual differences in model parameters (Heck et al., 2018). Küpper-Tetzel and Erdfelder (2012) have previously tested and validated the model successfully. Both an illustration of the model and the 12 model equations are available in the Electronic Supplemental Material (ESM) 1.

For our purposes, maintenance (m) and retrieval parameters (r_f) are key parameters to test the consolidation and the interference hypothesis, respectively. As detailed above, the interference hypothesis predicts that retrograde

facilitation results from enhanced accessibility of learned information in the testing phase. In the EMR model, this process is represented by the retrieval parameter r_f , the probability of successful retrieval of stored associations in free recall. In contrast, parameter r_c represents the probability of successful retrieval of the second word when the first word of an association stored in memory is provided as a cue. Given that the word pair is stored in memory, this cued recall probability should generally be very close to 1 in all conditions, as was previously found by Küpper-Tetzel and Erdfelder (2012). Parameter r_c is thus irrelevant for testing our hypotheses.

The consolidation hypothesis predicts that retrograde facilitation emerges as a consequence of enhanced maintenance across the retention interval. This process is represented by the maintenance parameters m_s and m_u . Küpper-Tetzel and Erdfelder (2012) observed that m_s and m_u can be equated to a single maintenance parameter m . Accordingly, the consolidation hypothesis predicts that m is larger in the alcohol condition compared to the placebo condition.

Parameter e should not be affected by alcohol administration after learning as it represents successful encoding of an association, a process that takes place before the experimental manipulation. Finally, parameters s and u represent probabilities of successful single word maintenance and retrieval for associations stored versus not stored in memory, respectively. They are thus not informative with respect to the question whether alcohol specifically enhances associative maintenance or retrieval.

It is important that the two parameters of prime interest, m and r_f , are not influenced by factors other than the administration of alcohol versus a placebo. It thus needs to be assured that (a) participants are completely sober when memory is tested after the retention interval to avoid contaminations of r_f with state-dependent learning decrements and (b) the retention interval does not include a night of sleep or daytime naps to preclude contaminations of successful maintenance (m) and retrieval (r_f) with sleep-related differences between conditions. For this purpose, we largely followed the original study by Parker et al. (1981) and employed a retention interval of 7 h that

ensured a sober state at both study and test. Additionally, the procedure was conducted in the controlled environment of a laboratory to preclude sleep, further consumption of alcohol, and other confounding variables. Note, however, that several methodological differences between our study and the original Parker et al. (1981) study remain. First, we did not restrict cognitive activity during the retention interval in any way (apart from enforcing compliance with the instructions). According to the interference hypothesis, the more cognitive activity during retention, the more accessibility benefits should emerge in the alcohol condition. Thus, a fair test of the interference hypothesis requires considerable cognitive activity in both conditions. Second, we used cued and free recall tasks for paired associates rather than assessing recognition memory for pictures. This adjustment is required by the EMR model. The same holds for the initial cued recall test immediately following learning, a procedure that deviates from Parker et al. (1981) but is mandatory for the EMR model-based analysis. Third, alcohol consumption was manipulated as a between-subjects factor in our study. We feel that this prevents alcohol-related expectancy effects more effectively than a within-subjects manipulation because participants lack a direct reference standard for the alcohol content. Fourth, we did not restrict our study to male participants but investigated all genders. We took into account the difference in alcohol tolerance between genders by administering different doses to women (0.51 g/kg) and men (0.59 g/kg).

Hypotheses

Because the retrograde facilitation effect emerged in all prior studies we were aware of, even in those in which its strength was probably hampered by state-dependent learning influences, we assumed that the effect truly exists and thus hypothesized that forgetting in cued recall across the retention interval (i.e., the difference in correct responses between the initial and the final cued recall) would be reduced for participants in the alcohol condition compared to the placebo condition (Hypothesis 1) and, in addition, that participants in the alcohol condition would retrieve more word associations in final free recall (Hypothesis 2).

Moreover, in line with the interference hypothesis, we hypothesized that the EMR model probability of retrieving word associations in free recall (parameter r_f) would be significantly higher for participants in the alcohol condition (Hypothesis 3). Based on the consolidation hypothesis, we additionally hypothesized that maintenance of stored word associations across the retention interval (parameter m) would be significantly higher for participants in the

alcohol condition (Hypothesis 4). No specific predictions applied to other model parameters.

Method

All methodological details of the present study were planned and evaluated by the authors in accordance with the ethical principles outlined in the Declaration of Helsinki (2013). The research protocol was approved by the ethics committee of the University of Mannheim. The materials necessary to replicate this study are available on the OSF (<https://doi.org/10.17605/OSF.IO/8E9PW>).

Design

The present study was conducted in a double-blind, randomized, placebo-controlled manner. Alcohol administration was a between-subjects factor. Thus, one group of participants consumed an individually determined dose of alcohol after learning while another group of participants received a perceptually indistinguishable placebo beverage. Participants were randomly assigned to one of the two conditions.

Participants

Sample Size

We defined a medium effect size of $d = 0.50$ as the minimum effect of interest in the present study. With $d = 0.50$, $\alpha = .05$, and a desired statistical power of $1 - \beta = .80$, a conventional Neyman-Pearson power analysis using the software G*Power (Faul et al., 2007) results in a required sample size of $N = 102$ for a one-tailed (i.e., directed) 2-groups t test (51 per experimental condition). To maximize efficiency, a sequential t test (e.g., Schnuerch & Erdfelder, 2020) was used to test the differences in cued and free recall memory performance between conditions, an approach that has been shown to reduce the required sample size to about 60% of the corresponding Neyman-Pearson sample size on average (see below). If both sequential tests would terminate with $n < 30$ in either condition, additional participants would be sampled subsequently until the threshold $n = 30$ would have been reached in each condition to enable meaningful MPT analyses of the data.

Recruitment

Potential participants were recruited via the online platform for study participation of the University of Mannheim and by advertisement within the university campus. Participants received study credit. The top three

performances in the immediate cued recall and the final free-then-cued-recall were rewarded with 20 € each. Study information advised subjects to participate in groups because they would watch movies, eat pizza, and drink alcohol as part of a study taking 8.5 hrs in total and allegedly investigating processing of movies under the influence of alcohol.

Eligibility

Interested individuals were required to report at least one heavy drinking episode (five alcoholic beverages for men and four alcoholic beverages for women at one occasion; Wechsler et al., 1995) within the last month to ensure that they had sufficient experience to tolerate the to-be-consumed alcohol dose. Additional inclusion criteria included age between 18 and 29 years, body mass index (BMI) between 18.5 and 29.9, no current or past diagnosis of a substance use disorder or any other addictive disorder, no current diagnosis of any other psychiatric disorder, no physical disorder that precludes the consumption of alcohol, no medical advice to avoid the consumption of alcohol, no current intake of any medication other than birth control, no pregnancy or possibility of pregnancy, and no lactation. Age and BMI restrictions were included to allow precise estimation of blood alcohol concentrations (BACs) by the Widmark formula (Thierauf et al., 2013). All criteria were assessed by self-report. Social drinking and no substance use disorder were further confirmed through the Alcohol Use Disorders Identification Test (AUDIT; see below).

Materials

Alcohol and Placebo

For participants assigned to the alcohol condition, the alcohol dose was 0.51 g/kg for women and 0.59 g/kg for men.¹ Based on the Widmark formula (Widmark, 1932, as cited in Thierauf et al., 2013) and following the refined procedure outlined by Thierauf et al. (2013), these doses could be expected to result in BACs of around 0.60‰ (see ESM 2), an intoxication level high enough to cause significant cognitive impairment. Assuming an elimination rate of around 0.15‰ per hour (Thierauf et al., 2013), it could be expected that alcohol would be eliminated from the body within the retention interval of 7 h. Moreover, the doses lie within the range of dosages that have successfully

been employed in previous research on retrograde facilitation, ranging from 0.40 g/kg (Parker et al. 1981) to 0.80 g/kg (e.g., Weafer et al., 2016a). The difference between the two doses for women and men accounts for the fact that women typically exhibit higher BACs than men for a given amount of consumed alcohol (Fillmore, 2001).

In the alcohol condition, a mixture of vodka, tonic water, and Tabasco sauce was served as described by Knowles and Duka (2004). The beverage in the alcohol condition was made up of one part vodka (40% alcohol content) and three parts tonic water such that the total amount of alcohol reached the individually predetermined dose. For administration, each individual beverage was split into 10 portions, meaning that every participant was asked to consume their respective total beverage in 10 smaller portions. To mask the taste and burn of vodka, two drops of Tabasco sauce were added to each portion. In the placebo condition, vodka was replaced with additional tonic water. In both conditions, glass rims were swabbed with vodka to further intensify the sensual impression of alcohol in the placebo condition (Lamberty et al., 1990). Hence, participants in the placebo condition actually did consume some alcohol, but a very low and negligible dose. Due to the inclusion criterion of a BMI between 18.5 and 29.9, the total amount of individual beverages in the alcohol condition could range between 252 ml (67 ml vodka + 185 ml tonic water, split into 10 portions of 25.2 ml each) in case of female gender and body weight of 42 kg, and 832 ml (221 ml vodka + 611 ml tonic water, split into 10 portions of 83.2 ml each) in case of male gender and body weight of 120 kg. Note that these values are theoretical extremes that were not encountered in our study. Actual beverages varied between 266 ml (71 ml vodka + 195 ml tonic water, split into 10 portions of 26.6 ml each) and 624 ml (166 ml vodka + 458 ml tonic water, split into 10 portions of 62.4 ml each).

Word Pairs

Forty German word pairs were taken from Hager and Hasselhorn (1994), the same as used by Küpper-Tetzel and Erdfelder (2012). These word pairs have the important characteristic of being only weakly associated to another. Hence, the probability of successful generation of the target word by mere guessing when presented with the cue word in cued recall tasks is minimized. A complete overview of the learning material is provided in ESM 3.

¹ Due to an initial error in calculating alcohol doses, the doses administered in this study were slightly lower than preregistered for the first $n_1 = 28$ participants ($N_{\text{total}} = 93$), i.e., approximately 0.44 g/kg for women and 0.51 g/kg for men. For the remaining $n_2 = 65$ participants, the preregistered doses were administered. Crucially, this difference in doses did not result in a relevant difference in peak BAC measurements in the alcohol condition, $M_1 = 0.42\text{‰}$ ($SD = 0.08$), $M_2 = 0.44\text{‰}$ ($SD = 0.08$), $t(44) = 0.60$, $p = .275$, Cohen's $d = 0.19$. Moreover, the lower-than-preregistered doses still lie above the medium dose of 0.40 g/kg in Parker et al. (1981), which was sufficient to yield a substantial retrograde facilitation effect.

AUDIT

A German translation of the Alcohol Use Disorders Identification Test (AUDIT; Saunders et al., 1993) was used to check the exclusion criterion of hazardous drinking habits and the inclusion criterion of social drinking. This scale is recommended in German clinical guidelines for the screening of alcohol use disorders (K. Mann et al., 2017). It consists of 10 items asking for the amount and frequency of drinking occasions and the occurrence of several negative consequences following drinking. Items include three to five alternatives that are rated with a score between 0 and 4. Thus, the maximum total score is 40. Total scores of 8 or more are treated as indicators of harmful alcohol use (Babor et al., 2001). Therefore, interested persons scoring above 7 were not allowed to participate. The inclusion criterion of social drinking was ensured by allowing participation only if individuals reached a total score of at least 2 on the first two items (i.e., one drinking occasion per month with at least three to four alcoholic drinks or two to four drinking occasions per month with at least one or two alcoholic drinks) and a score of at least 1 on the third item (i.e., heavy drinking episodes in the past).

Procedure

Individuals interested in participating in this study first accessed an online survey that provided them with a description of the study procedure. To prevent demand effects, participants were told that (a) all participants in this study would consume alcohol in a more or less high dose and that (b) the focus of this study lay on the perception and processing of movies under the influence of alcohol. Inclusion criteria were also checked at this occasion. Eligible participants were asked not to drink alcohol within 48 hrs prior to their participation in the experiment and to have lunch before participation began. The experiment took place in a classroom-like laboratory within the premises of the Department of Psychology at the University of Mannheim.

Participants were asked to arrive at 1:30 p.m. at the laboratory. With the exception of general information and instructions regarding the procedure, the first phase of the experiment was administered individually. First, participants were asked for inclusion criteria again and provided written consent. Additionally, breath alcohol concentration was measured to ensure soberness of all participants. Then, participants were instructed for the learning phase. Forty word pairs were presented in randomized order for 5 s each on a computer screen, with participants being informed that memory for the materials would be tested afterward in a cued recall task. The announced immediate

cued recall was conducted following a short distractor task where participants were asked to sequentially assess the correctness of 15 equations as quickly as possible. In the immediate cued recall, participants were required to complete each cue word with the respective target word. Cue words were presented in randomized order on a computer screen with participants responding by filling in the respective target words at their own pace. After completion, participants were asked to drink 10 portions of a beverage (either the alcohol or the placebo beverage, depending on the condition they were assigned to) within 30 min (i.e., 3 min per portion). Both the experimenter and the participant were blind with respect to the experimental condition. Following the first portion, participants were asked to estimate the alcohol content of the beverage on a scale from “less than 1%” to “equal to or more than 20%” (note that the true value was about 10% for all participants in the alcohol condition) to check whether the placebo beverage successfully blinded participants for their experimental condition (Keane et al., 1980). After consuming the final portion, participants of both conditions were asked to rinse their mouth with water. Participants were instructed individually not to communicate with other participants about their experiences during learning and alcohol administration. This completed the first phase of this study that lasted approximately 1 h.

The retention interval of 7 h was spent by participants in a seminar room under permanent supervision. Because participants were encouraged to register in groups, they could freely interact and socialize with their colleagues during this time. As part of the cover story, movies were presented. As participation in the study covered the evening hours, pizza was provided for dinner. Additionally, nonalcoholic drinks were available all the time. Breath alcohol concentrations of all participants were checked 30, 60, and 90 min after the administration of alcohol or a placebo (to assess peak BACs) and immediately before memory testing began (to ensure soberness of all participants). Participants were told to abstain from drinking alcohol and sleeping over the whole course of the retention interval. Compliance with the instruction not to communicate about the learning phase and the alcohol administration was monitored by the experimenter.

After the retention interval, participants performed a surprise free-then-cued-recall for the learned word associations. In the free recall task, participants were asked to recall as many of the word pairs as possible within 8 min. Importantly, participants were instructed to write down single words in cases when they did not remember both words of an association. Next, a final cued recall identical to the immediate cued recall task after the study phase was employed. Additionally, participants were asked to declare whether they followed the instruction not to communicate

about their experiences from the first phase of the experiment and whether they thought about or actively rehearsed any word pairs in a postexperimental questionnaire. Finally, participants were informed about the true theoretical background and design of the study and the exact dose of alcohol they consumed. Additionally, research assistants were present to answer further questions and concerns. The study ended by 10.00 p.m.

Data Analysis

All statistical tests were conducted with $\alpha = .05$. As a manipulation check, peak BACs in the alcohol condition were expected to be significantly larger than zero and significantly larger than BACs observed in the placebo condition. BACs in the placebo condition were expected never to rise to a level larger than zero. Blinding of participants for their condition would be regarded as successful if participants in the placebo condition estimated the alcohol content of their beverage as 1% or more, indicated by all responses other than “less than 1%”. Analyses were conducted with and without those participants in the placebo condition who estimated the alcohol content as less than 1% to test whether results were affected by BAC awareness. Similarly, analyses were conducted with and without those participants who communicated about Phase 1 of the experiment with other participants.

To maximize efficiency of data collection, the sampling process was conducted using a one-tailed group-sequential probability ratio (SPRT) t test. Thus, after each group of participants investigated as described above, the resulting data were analyzed to decide whether a decision for or against Hypothesis 1 could be made or further data were required. This sequential procedure has been shown to result in a considerable reduction of required sample sizes when compared to conventional Neyman-Pearson t tests. At the same time, error probabilities are controlled (Schnuerch & Erdfelder, 2020). In our application, standard parameters for the Type-1 and Type-2 error probabilities were used (i.e., $\alpha = .05$ and $\beta = .20$, respectively, assuming a medium effect size $d = .50$, cf. Cohen, 1988). Difference scores were calculated for every participant, subtracting the number of correct responses in the final cued recall from the number of correct responses in the immediate cued recall. Then, the difference between the two mean difference scores of experimental conditions were tested for statistical significance using the SPRT t test as specified above until a decision was made.

To test Hypothesis 2, the number of complete word pairs recalled was calculated for every participant to obtain mean performances for both experimental conditions. The difference of these two means was tested for statistical

significance with the same group-sequential one-tailed two-sample SPRT t test as in case of Hypothesis 1.

Hypotheses 3 and 4 referred to parameters r_f and m in the EMR model. Participants underperforming (<30% correct) or overperforming severely (>80% correct) in the immediate cued recall needed to be excluded from MPT analyses to ensure sufficient data points for event categories E_1 to E_6 and E_7 to E_{12} , respectively. A minimum sample size of $n = 30$ per condition was ensured to enable trustworthy parameter estimates for the EMR model in either of the conditions. Next, the frequencies of the 12 event categories E_1 to E_{12} were calculated individually for all remaining participants and aggregated within conditions. The resulting data could then be used to fit the EMR model (see above). All MPT analyses were conducted twice using TreeBUGS (Heck et al., 2018) for individual data and multiTree (Moshagen, 2010) for aggregated data. Following Küpper-Tetzl and Erdfelder (2012), we would (a) try to equate m_s and m_u so that a single parameter m represents the probability of successful maintenance and (b) equate r_c across experimental conditions. For the aggregated data, this leads to 2 (= number of conditions) $\times 5$ (= unrestricted model parameters) $+ 1$ (= number of parameters equated for both conditions) = 11 model parameters to be estimated given 2 (= number of conditions) $\times 11$ (= number of free categories per condition) = 22 independent category frequencies. Accordingly, the number of degrees of freedom of the G^2 goodness-of-fit tests test is $22 - 11 = 11$. We would accept the model fit if $G^2(11)$ would result in a p value greater than $.05$. If this criterion would not be met, the model would need to be replaced by a more general model version, for example, by allowing m , s , and/or u to differ between successful and unsuccessful immediate cued recall.

To test Hypothesis 3, an equality constraint would be imposed on the r_f parameters of both experimental conditions. If the resulting decrease in model fit $\Delta G^2(1)$ would become significant at $\alpha = .05$ (one-tailed) and parameter estimates would be in the direction predicted by the interference hypothesis, the latter hypothesis would be confirmed. For the test of Hypothesis 4, analogous procedures would be used: An equality constraint would be imposed on the m parameters of both experimental conditions, and the hypothesis would be confirmed if the resulting decrease in model fit $\Delta G^2(1)$ would become significant at $\alpha = .05$ (one-tailed) and parameter estimates would be in the predicted direction.

Results

The full data set used for the analyses and the corresponding codebook are provided on the OSF (<https://doi.org/10.1027/1618-3169/a000569>).

org/10.17605/OSF.IO/2K4J7). Additionally, the R code to obtain the results reported below is provided in ESM 4.

Sample

A total of $N = 93$ participants (divided into 20 sequential groups, 2–10 participants per group, 46 participants in the alcohol condition) took part in the experiment. Although a decision for both SPRT t tests was already reached earlier (see below), it was not before this point of the data collection process that the preregistered minimum of $n = 30$ participants per condition was achieved for our encoding-maintenance-retrieval MPT analyses. In the full sample, 62 participants were female, and 31 participants were male. The mean age was 20.67 years ($SD = 2.18$). A majority of 91 participants reported to be university students, of which 72 participants were enrolled in a psychology program. The mean AUDIT score was 5.39 ($SD = 1.23$).

Control Variables and Manipulation Checks

BAC measurements at the beginning of the experiment confirmed that all participants arrived at the laboratory completely sober. Six participants reported to have consumed alcohol within 48 h before the experiment. Most participants in the alcohol condition (41 of 46 participants) reached their measured peak BAC 30 min after the end of the alcohol administration, four participants reached it after 60 min, and one participant after 90 min. The mean peak BAC in the alcohol condition was 0.43‰ ($SD = 0.08$, Range = 0.22–0.62). At all three measurement occasions after the alcohol administration, mean BACs in the alcohol condition were significantly larger than zero: After 30 min, $M_{BAC} = 0.43‰$ ($SD = 0.08$, Range = 0.22–0.62), $t(45) = 34.74$, $p < .001$; after 60 min, $M_{BAC} = 0.37‰$ ($SD = 0.08$, Range = 0.18–0.57), $t(45) = 31.40$, $p < .001$; and after 90 min, $M_{BAC} = 0.30‰$ ($SD = 0.08$, Range = 0.12–0.46), $t(42) = 24.91$, $p < .001$.² Eight of 47 participants in the placebo condition estimated the alcohol content of their beverage to be below 1%. Actual BACs in the placebo condition never rose to a level larger than zero, except for one participant who had a BAC measure of 0.08‰ 90 min after the end of the placebo administration. All participants in both conditions were completely sober again at the last BAC measurement immediately before the final recall tests.

In the postexperimental questionnaire, 23 participants (13 from the alcohol condition, 10 from the placebo

condition) reported to have communicated about their beverages, and two participants (one from the alcohol condition, one from the placebo condition) reported to have communicated about both the beverages and the word pairs during the retention interval. Thirty-three participants (14 from the alcohol condition, 19 from the placebo condition) reported to have thought about the word pairs during the retention interval. Of these, three participants (two from the alcohol condition, one from the placebo condition) reported to have engaged in active rehearsal of the word pairs.

Design-Based Results

Hypotheses 1 and 2 referred to the difference in correct responses between immediate and final cued recall and the number of complete word pairs reproduced during free recall, respectively. Both hypotheses were tested using the SPRT t test (Schnuerch & Erdfelder, 2020) implemented in the `sprtt` R package (v0.1.0; Steinhilber et al., 2021). For both Hypotheses 1 and 2, the decision to accept the null hypothesis of no significant alcohol benefit was reached rather quickly after $N = 18$ subjects (divided into five sequential groups, $n_{\text{Alcohol}} = n_{\text{Placebo}} = 9$) had participated. The means and standard deviations from this subsample and the full sample ($N = 93$) for the immediate cued recall, the final cued recall, the cued recall difference (Hypothesis 1), the number of complete word pairs reproduced during free recall (Hypothesis 2), and the number of single words reproduced during free recall are reported in Table 2.

For Hypothesis 1, the likelihood ratio (LR) of the SPRT t test at $N = 18$ was $LR_{18} = 0.18$, thereby undercutting the lower SPRT threshold implied by our preregistered SPRT parameters, that is, $\beta/(1 - \alpha) = 0.20/0.95 = 0.21$. The observed LR indicates that the data at $N = 18$ were about $1/0.18 = 5.6$ times more likely under H_0 than under H_1 , thus enforcing acceptance of H_0 . The sample estimate of Cohen's d was -0.53 , that is, there was a medium-sized effect in the direction opposite to Hypothesis 1. For Hypothesis 2, we observed $LR_{18} = 0.11$. Thus, at $N = 18$, the data were about $1/0.11 = 9.1$ times more likely under H_0 than under H_1 , also enforcing acceptance of H_0 . In this case, there was a strong effect in the direction opposite to Hypothesis 2, Cohen's $d = -0.82$. In sum, no significant retrograde facilitation effect could be observed for either of our two dependent variables. Plots of the developments of the log-likelihood ratios for both SPRT t tests are provided in ESM 5.

² Due to technical problems during the experiment, BAC measurements are missing for three participants in the alcohol condition 90 min after the end of the alcohol administration and immediately before the final recall tests.

Table 2. Means and standard deviations (SD) of all dependent variables in the alcohol and placebo condition

Dependent variable	SPRT subsample ($N = 18$)		Full sample ($N = 93$)	
	Alcohol	Placebo	Alcohol	Placebo
Immediate cued recall	19.33 (7.94)	26.56 (4.64)	23.57 (8.89)	23.83 (7.50)
Final cued recall	16.78 (8.80)	24.89 (4.83)	21.85 (9.59)	21.34 (7.65)
Cued recall difference	2.56 (1.81)	1.67 (1.50)	1.72 (1.73)	2.49 (2.41)
Free recall: complete pairs	9.22 (4.60)	13.22 (5.09)	11.57 (5.44)	10.04 (4.66)
Free recall: single words	23.67 (8.89)	31.56 (9.89)	28.11 (10.68)	26.21 (9.29)

Note. Decisions in favor of H_0 for Hypotheses 1 (cued recall difference) and 2 (free recall: complete pairs) were reached by the sequential probability ratio t tests (SPRT t tests) after data from $N = 18$ participants had been collected. Due to the second stopping criterion of $n = 30$ in both conditions for the multinomial processing tree (MPT) analysis (after excluding participants over- or underperforming in the immediate cued recall), a total of $N = 93$ subjects participated in this study. *SDs* in parentheses.

Although irrelevant for the SPRT t test decisions in favor of H_0 , we additionally inspected the likelihood ratios of the full-sample SPRT t tests, LR_{93} , for explorative reasons. Moreover, because LR_{93} assumes a fixed effect size of $d = 0.50$ under H_1 – a debatable assumption – we additionally computed corresponding Bayes factors (BF_{10}) using the default Cauchy prior as implemented in the BayesFactor R package (v0.9.12-4.2; Morey & Rouder, 2018). As summarized in Table 2, the descriptive pattern regarding our two dependent variables of interest was reversed in the full sample of $N = 93$ as compared to the subsample of $N = 18$ required until termination of the SPRT. However, likelihood ratios in the full sample indicate no clear evidence in favor of H_1 in both cases; $LR_{93} = 3.84$, Cohen's $d = 0.37$ for Hypothesis 1 and $LR_{93} = 1.81$, Cohen's $d = 0.30$ for Hypothesis 2. In line with this, Bayesian t tests revealed $BF_{10} = 1.63$ for Hypothesis 1 and $BF_{10} = 1.00$ for Hypothesis 2. Thus, although Hypotheses 1 and 2 are descriptively in line with the data of the full sample, the evidence in favor of H_1 is negligible in both cases (Kass & Raftery, 1995).

To further scrutinize the robustness of our conclusions regarding Hypotheses 1 and 2, we excluded participants from both the subsample and the full sample who either (a) were assigned to the placebo condition and estimated the alcohol content of their beverage to be less than 1% or (b) reported to have communicated about Phase 1 of the experiment with other participants. The means and standard deviations for the two dependent variables of interest are reported in Table 3.

After the application of both exclusion criteria, $N = 12$ participants remained in the subsample, whereas $N = 62$ participants remained in the full sample. The descriptive patterns mirror those obtained without consideration of these exclusion criteria. Similarly, corresponding likelihood ratios and Bayes factors support the conclusion of no clear evidence in favor of H_1 : For Hypothesis 1, $LR_{12} = 0.39$, $LR_{62} = 2.54$, $BF_{10} = 1.15$; for Hypothesis 2, $LR_{12} = 0.26$, $LR_{62} = 1.53$, $BF_{10} = 0.81$.

Model-Based Results

Hypotheses 3 and 4 referred to the retrieval parameter r_f and the maintenance parameter m of the EMR model, respectively. For the aggregated data, we conducted all analyses in multiTree (Moshagen, 2010). The corresponding preregistered analyses are provided in ESM 6. The baseline EMR model as described in the Data Analysis section did not meet our model fit criterion with $\alpha = .05$, $G^2(11) = 22.00$, $p = .024$, $AIC = 9,758.71$, $BIC = 9,824.18$. Therefore, we defined a generalized model version with parameters m , u , and s allowed to vary within both conditions between successful and unsuccessful immediate cued recall (m_s and m_u , u_s and u_u , and s_s and s_u , respectively). This generalized model fit the data well, $G^2(5) = 5.97$, $p = .309$, $AIC = 9,754.68$, $BIC = 9,855.85$. A close inspection of this model revealed a significant difference between parameters u_s and u_u only within the placebo condition, $\Delta G^2(1) = 12.55$, $p < .001$, but not within the alcohol condition, $\Delta G^2(1) = 1.06$, $p = .303$. Thus, participants in the placebo condition had a significantly higher probability of single word retrieval in the free recall for unsuccessfully stored associations after successful immediate cued recall ($u_s = .21$) than after unsuccessful immediate cued recall ($u_u = .10$), in contrast to participants in the alcohol condition ($u_s = .16$, $u_u = .12$).

Accordingly, we defined a new baseline model with the following restrictions: In line with Küpper-Tetzl and Erdfelder (2012), parameters m_s and m_u as well as s_s and s_u were equated within both experimental conditions, and parameter r_c was equated across conditions. In contrast, parameter u was allowed to vary freely within both conditions between successful (u_s) and unsuccessful (u_u) immediate cued recall. This redefined model yielded a good fit to the data, $G^2(9) = 7.43$, $p = .592$, $AIC = 9,748.14$, $BIC = 9,825.51$. As this model version is more parsimonious than the generalized model version, we decided to use it as our baseline model for the hypothesis tests.

Table 3. Means and standard deviations (SD) of both dependent variables of interest in the alcohol and placebo conditions after exclusion criteria were applied

Dependent variable	SPRT subsample ($N = 12$)		Full sample ($N = 62$)	
	Alcohol	Placebo	Alcohol	Placebo
Cued recall difference	2.67 (2.07)	2.00 (1.26)	1.91 (1.86)	2.73 (2.57)
Free recall: complete pairs	9.67 (4.84)	13.33 (5.75)	11.47 (5.97)	9.80 (4.92)

Note. Participants from the crucial sequential probability ratio test (SPRT) subsample and the full sample were excluded if they estimated the alcohol content of their beverage to be less than 1% in the placebo condition and if they reported to have communicated about Phase 1 of the experiment with other participants. SDs in parentheses.

This baseline model was additionally fitted to the individual data in the form of a Bayesian hierarchical model. For this purpose, the model was estimated for both conditions separately, using the latent trait framework of Klauer (2010) as implemented in TreeBUGS (Heck et al., 2018). In this Bayesian framework, convergence of parameter estimates can be evaluated by means of the potential scale reduction factor R (Gelman & Rubin, 1992). Good convergence was obtained for all parameters in the alcohol condition, $\hat{R} \leq 1.007$, and also in the placebo condition, $\hat{R} \leq 1.003$. The model fit was evaluated using the goodness-of-fit statistics T_1 and T_2 (Klauer, 2010). Good model fit was obtained for either condition, as indicated by posterior predictive p values of $p_1 = .288$ ($p_2 = .300$) in the alcohol condition and $p_1 = .565$ ($p_2 = .420$) in the placebo condition. In Bayesian hierarchical MPT modeling, statistical reliability of parameter differences is typically assessed by checking (a) whether the 95% Bayesian credibility interval (BCI) of the posterior distribution of the difference estimate does not include zero (two-tailed) or (b) whether the Bayesian p value, that is, the proportion of the posterior distribution of the difference estimate below zero is smaller than the chosen significance level of $\alpha = .05$ (one-tailed).

After excluding all participants who underperformed ($<30\%$ correct) or overperformed severely ($>80\%$ correct) in the immediate cued recall, a subsample of $N = 71$ participants remained for the preregistered model-based analysis (31 in the alcohol condition, 40 in the placebo condition). The parameter estimates for both the aggregated and the individual data are presented in Table 4.

As expected, the probability of associative encoding (parameter e) did not differ significantly between conditions, neither for the aggregated data, $\Delta G^2(1) = 0.55$, $p = .460$, nor for the individual data, 95% BCI = $[-.09, .06]$. Additionally, as expected, parameter r_c was estimated to be very close to 1 in both modeling approaches.

Based on the interference hypothesis, we predicted a significantly higher probability of associative retrieval in free recall (parameter r_f) for participants in the alcohol condition (Hypothesis 3). Parameter estimates of both modeling approaches were in line with this prediction, and this descriptive effect did reach statistical significance (one-tailed) both for

the aggregated data, $z = \sqrt{\Delta G^2(1)} = 2.35$, $p = .009$, and for the individual data, Bayesian $p = .023$.

Based on the consolidation hypothesis, we hypothesized the probability of associative maintenance (parameter m) to be significantly higher for participants in the alcohol condition (Hypothesis 4). Again, parameter estimates of both modeling approaches were descriptively in line with our prediction. However, this descriptive effect was statistically significant (one-tailed) only for the aggregated data, $z = \sqrt{\Delta G^2(1)} = 2.33$, $p = .010$, but not for the individual data, Bayesian $p = .104$.

Overall, the results of our preregistered MPT analysis were in line with Hypothesis 3, but did not yield clear evidence in favor of Hypothesis 4. Additionally, the descriptive effect for parameter m was rather small, irrespective of the modeling approach. To further clarify the robustness of this conclusion, we conducted an exploratory MPT multiverse analysis. In this analysis, we not only included our preregistered exclusion criteria of immediate cued recall performance, alcohol content estimation in the placebo condition, and communication during the retention interval (see the “Data Analysis” section), but additionally considered the influence of failure not to drink alcohol within 48 h prior to the experiment and active word pair rehearsal during the retention interval. Thus, our multiverse analysis had a 2 (modeling approach: aggregated vs. individual data) \times 2 (model-related exclusion criterion present vs. absent: under- or overperforming in the immediate cued recall) \times 2 (at least one procedure-related exclusion criterion present vs. absent: drinking alcohol within 48 h prior to the experiment, estimating the alcohol content to be below 1% in the placebo condition, communicating about the alcohol vs. placebo administration and/or the learning phase, actively rehearsing word pairs during the retention interval) design.

The first two cells of this multiverse design contained our preregistered analyses reported above, where both modeling approaches were applied to the data of all participants who did not meet the model-related exclusion criterion, whereas procedure-related exclusion criteria were not considered. The results of the remaining cells are reported in ESM 7. For both the aggregated and the individual data, the model fit the data well in all remaining

Table 4. Parameter estimates from the main multinomial processing tree (MPT) analysis

Parameter	Alcohol condition				Placebo condition			
	Aggregated data		Individual data		Aggregated data		Individual data	
	MLE	95% CI	<i>M</i>	95% BCI	MLE	95% CI	<i>M</i>	95% BCI
<i>e</i>	.58	[.55, .61]	.59	[.53, .64]	.60	[.57, .62]	.60	[.55, .65]
<i>m</i>	.92	[.90, .95]	.93	[.90, .96]	.89	[.86, .91]	.90	[.86, .93]
<i>r_c</i>	.98	[.97, .98]	.98	[.97, .99]	.98	[.97, .98]	.98	[.97, .99]
<i>r_f</i>	.50	[.46, .54]	.51	[.45, .56]	.44	[.41, .47]	.43	[.39, .48]
<i>s</i>	.05	[.03, .07]	.05	[.03, .07]	.11	[.09, .13]	.11	[.09, .14]
<i>u_s</i>	.16	[.08, .24]	.18	[.07, .34]	.22	[.15, .28]	.20	[.12, .28]
<i>u_u</i>	.12	[.10, .14]	.11	[.08, .15]	.10	[.09, .12]	.11	[.09, .13]

Note. For the aggregated data, the encoding-maintenance-retrieval (EMR) model was fitted using the multiTree software (Moshagen, 2010). Maximum likelihood parameter estimates (MLE) are presented alongside the corresponding 95% confidence interval (CI). Parameter r_c was equated across conditions. For the individual data, the model was fitted for both conditions separately using the R package TreeBUGS (Heck et al., 2018). Posterior means (*M*) are presented alongside the corresponding 95% Bayesian credibility intervals (BCI).

cells, thereby allowing for an interpretation of the respective parameter estimates. Overall, our conclusions for both Hypotheses 3 and 4 were confirmed: Parameter r_f differed reliably between conditions in all cells for the aggregated data (all $p \leq .034$) and in all cells except one (Bayesian $p = .051$) for the individual data (both other Bayesian $p \leq .016$). In contrast, parameter m differed reliably in only one additional cell ($p = .010$) for the aggregated data (both other $p \geq .090$) and in no cell for the individual data (all Bayesian $p \geq .132$).

Discussion

In the present study, we attempted to replicate the counterintuitive phenomenon of alcohol-induced retrograde facilitation. Using a retention interval of 7 hrs, we followed the basic procedure introduced by Parker et al. (1981) in their original study. Thus, we avoided possible confounds of interpolated sleep and other memory-relevant activities and thereby put the original observation of the effect of interest to a critical test. Moreover, we used a sequential testing procedure (i.e., the SPRT t test) to increase the efficiency of data collection and applied the encoding-maintenance-retrieval MPT model to allow for an assessment of the underlying mechanisms. Finally, we decided to conduct and report our study in a registered report format to make the severity of our hypothesis tests transparent (Lakens, 2019).

Contrary to our predictions, we found that drinking alcohol immediately after learning did not significantly increase performance on a delayed subsequent memory test. In line with this finding, there was no clear evidence for a latent maintenance benefit in the alcohol condition, neither in our preregistered MPT analysis nor in an

exploratory MPT multiverse analysis. Importantly, however, these MPT analyses revealed a reliable and robust retrieval benefit in the alcohol condition. Thus, although we failed to replicate the original finding by Parker et al. (1981) using conventional measures of memory performance (i.e., cued and free recall), we did find clear-cut evidence in favor of an underlying retrieval benefit (i.e., an effect on EMR parameter r_f).

When evaluating our failure to successfully replicate the results by Parker et al. (1981) on the manifest memory measures of cued and free recall, several methodological differences between the original study and our replication study need to be considered. Most importantly, the results from our MPT analyses suggest that observing a strong and significant retrograde facilitation effect in behavioral measures might require a study design that maximizes the latent contributions of retrieval processes to final memory performance, while minimizing encoding and maintenance contributions. In this respect, it is important to note that Parker et al. (1981) presented their participants with pictures, while we used word pairs as learning material. Crucially, the memory-system dependent forgetting hypothesis (Hardt et al., 2013; Kuhlmann et al., 2021) suggests that hippocampus-dependent memory traces should be less susceptible to retroactive interference than extra-hippocampally represented memories. Thus, the contribution of increased retrievability as a result of reduced retroactive interference might be less pronounced when using paired associates such as word pairs as learning material. Indeed, most other studies reported in the literature used lists of single items instead of paired associates as learning material to demonstrate retrograde facilitation by alcohol (e.g., novel words, Carlyle et al., 2017). On the other hand, Parker et al. (1981) found a significant retrograde facilitation effect on recognition performance, although recognition performance should

not or only minimally rely on the retrievability of information from memory.

In contrast to our replication study and the original Parker et al. (1981) study, most studies in the literature used a retention interval between alcohol versus placebo administration and final memory test that was considerably longer than 7 h, ranging between 16 and 48 h (Bruce & Pihl, 1997; Carlyle et al., 2017; Lamberty et al., 1990; Mueller et al., 1983; Parker et al., 1980, study 2; Weafer et al., 2016a, 2016b). As a consequence, these studies could not control for memory-relevant behaviors during the retention interval and necessarily included at least one night of sleep. Against the backdrop of our replication failure, the possibility that the strength of the retrograde facilitation effect was overestimated in these studies due to an effect of acute alcohol intoxication on the sleep architecture – and, thereby, on sleep-induced memory consolidation, potentially benefitting both maintenance and retrieval processes – cannot be dismissed.

Certain limitations of our current study should also be recognized. First, in line with common conventions (cf. Cohen, 1988), we realized a (preregistered) statistical power of $1 - \beta = .80$ and a minimum effect of interest of $d = .50$ for our SPRT t tests of Hypotheses 1 and 2. Thus, compared to false-positive findings (probability $\alpha = .05$), the probability of false-negative findings ($\beta = .20$) was relatively high in our study, especially if the true d happened to be less than $.50$. However, a higher statistical power (e.g., $1 - \beta = .95$) for smaller effect sizes (e.g., $d = .20$) would have increased the expected sample size considerably. Thus, we believe to have found a reasonable compromise between sufficient statistical power on the one hand and feasibility of a very resource-intensive study procedure on the other hand.

Second, both SPRT t tests terminated rather quickly such that a decision for both Hypotheses 1 and 2 could be made after only $N = 18$ subjects had participated in the study. This outcome nicely highlights the superiority of sequential probability ratio tests in terms of efficiency (cf. Erdfelder & Schnuerch, 2021; Schnuerch & Erdfelder, 2020; Schnuerch et al., 2022). However, in our specific case, the second stopping criterion required for the MPT analyses ($n \geq 30$ in both conditions after exclusion of participants under- or overperforming in the immediate cued recall) enforced continuation of data collection until a total of $N = 93$ subjects had participated. Thus, our statistical decisions regarding Hypotheses 1 and 2 were based on a fraction of only 19% of the full sample. Although this procedure is statistically valid (i.e., it satisfies the preregistered statistical error probabilities α and β) and empirically in line with Bayesian t test outcomes, we understand that basing hypothesis tests on such a small percentage of the full sample might seem unsatisfactory. A conceivable alternative to our preregistered strategy would

have been to start the SPRT t tests only when the MPT stopping criterion had already been met. However, although avoiding the aforementioned problem, the efficiency of the data collection process would have suffered from such an approach. In our view, this shows that best practice recommendations of how to optimally combine sequential testing procedures such as the SPRT t test with additional stopping rules are needed.

Third, measured peak BACs in the alcohol condition were somewhat lower than anticipated. While the alcohol doses of 0.51 g/kg for women and 0.59 g/kg for men were expected to result in peak BACs of about 0.60‰, the observed mean peak BAC was $M = 0.43$ ‰. Two factors might be considered to explain this discrepancy. First, peak BACs were most likely reached before or after our measurements for almost all participants, such that our measures underestimate the true peak BACs. Second, alcohol administration was based on the self-reported body weights of participants. It seems reasonable to assume that participants either underestimated and/or knowingly understated their true body weight due to reasons of social desirability. Crucially, the mean peak BAC measured in our study was still higher than in the medium (0.5 ml/kg) alcohol condition of Parker et al. (1981; 0.34‰) for which these authors reported a significant retrograde facilitation effect.

Despite these limitations, our results clearly cast doubt on the reliability, strength, and generalizability of the retrograde facilitation effect in behavioral memory measures. Importantly, the effect is viewed by some authors as a central piece of evidence for the theoretical claim that memory consolidation processes are not limited to periods of sleep, but rather occur whenever retroactive interference is reduced, such as during acute alcohol intoxication (Mednick et al., 2011; Wixted, 2004, 2010). Our lack of evidence for alcohol effects on the probability of maintaining paired associates in memory suggests that such a theoretical model of episodic memory should be treated with caution. Instead, we found that postencoding alcohol administration had a beneficial effect on the probability of retrieving word pairs from memory during free recall. This result is in line with the interference hypothesis, which suggests that retrograde facilitation emerges as a direct consequence of reduced retroactive interference. As such, our finding is more in line with theoretical models that attribute episodic memory improvements (e.g., sleep-induced retrograde facilitation; Rasch & Born, 2013) to retrieval benefits resulting from increased temporal distinctiveness of memory traces (Brown et al., 2007; Ecker et al., 2015).

In sum, our mixed pattern of results best resonates with the idea that alcohol-induced retrograde facilitation really exists but is limited to a retrieval benefit caused by retroactive interference reduction. This idea nicely accounts

not only for our confirmation of Hypothesis 3 but also for our failures to confirm Hypotheses 1 and 4 (as both hypotheses refer to memory measures more sensitive to storage than to retrieval from memory). From this theoretical perspective, the only inconsistent outcome is our rejection of Hypothesis 2 concerning free recall. A possible explanation is that overall free recall performance is a less pure measure of retrieval capacity than EMR parameter r_f which specifically captures retrieval of stored word pairs in free recall.

In terms of future research, more replication studies are needed to specify the exact conditions under which alcohol-induced retrograde facilitation can or cannot be observed. These studies should explicitly consider the role of the learning material (item vs. associative memory) and of sleep and other memory-relevant activities during the retention interval. This implies that more studies are needed that adopt the basic procedure used by Parker et al. (1981) to effectively control behaviors of participants during the retention interval. To make such resource-intensive studies feasible, they should be designed to amplify underlying retrieval differences between conditions (see above), and methodological and statistical innovations such as sequential testing procedures should be embraced. Additionally, as shown by the results of our study, the adoption of MPT modeling for hypothesis generation and testing can be expected to extend possible insights beyond the scope of more conventional analysis strategies. An a priori power analysis using multiTree (Moshagen, 2010) based on the parameter estimates from our pre-registered MPT analysis suggests a minimum total sample size of $N = 66$ across conditions (with 40 word pairs per participant) for future EMR studies that test alcohol versus placebo differences in maintenance and/or retrieval probabilities (condition difference in m or r_f under $H_1 = .10$, $\alpha = .05$, $1 - \beta = .95$).

Alcohol-induced retrograde facilitation is a fascinating psychological phenomenon with potentially far-reaching theoretical and practical implications. However, our results show that more research is clearly needed to provide a solid empirical basis for further discussions of such implications. We hope that future research will build on the methodological and statistical innovations applied in this study to arrive at a deeper understanding of the moderators and mediators of retrograde facilitation by alcohol.

Electronic Supplementary Material

The electronic supplementary material is available with the online version of the article at <https://doi.org/10.1027/1618-3169/a000569>

ESM 1. This text file provides an illustration of the EMR model and the corresponding model equations.

ESM 2. This text file details the calculation of alcohol doses as a function of gender and body weight.

ESM 3. This text file provides a complete overview of the learning material.

ESM 4. This R script contains all analysis steps to obtain the reported results (except MPT results for aggregated data).

ESM 5. This text file depicts the developments of the log-likelihood ratios of both SPRT t test.

ESM 6. This multiTree file contains all analysis steps to obtain the reported MPT results for aggregated data.

ESM 7. This text file contains the results of the MPT multiverse analysis.

References

- Babor, T. F., Higgins-Biddle, J. C., Saunders, J. B., & Monteiro, M. G. (2001). *AUDIT: The alcohol use disorders identification test: Guidelines for use in primary care* (2nd ed.). World Health Organization.
- Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin & Review*, 6(1), 57–86. <https://doi.org/10.3758/BF03210812>
- Brown, G. D. A., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review*, 114(3), 539–576. <https://doi.org/10.1037/0033-295X.114.3.539>
- Bruce, K. R., & Pihl, R. O. (1997). Forget “drinking to forget”: Enhanced consolidation of emotionally charged memory by alcohol. *Experimental and Clinical Psychopharmacology*, 5(3), 242–250. <https://doi.org/10.1037/1064-1297.5.3.242>
- Carlyle, M., Dumay, N., Roberts, K., McAndrew, A., Stevens, T., Lawn, W., & Morgan, C. J. A. (2017). Improved memory for information learnt before alcohol use in social drinkers tested in a naturalistic setting. *Scientific Reports*, 7(1), Article 6213. <https://doi.org/10.1038/s41598-017-06305-w>
- Chan, J. K. M., Trinder, J., Andrewes, H. E., Colrain, I. M., & Nicholas, C. L. (2013). The acute effects of alcohol on sleep architecture in late adolescence. *Alcoholism: Clinical and Experimental Research*, 37(10), 1720–1728. <https://doi.org/10.1111/acer.12141>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge. <https://doi.org/10.4324/9780203771587>
- Ebrahim, I. O., Shapiro, C. M., Williams, A. J., & Fenwick, P. B. (2013). Alcohol and sleep I: Effects on normal sleep. *Alcoholism: Clinical and Experimental Research*, 37(4), 539–549. <https://doi.org/10.1111/acer.12006>
- Ecker, U. K. H., Brown, G. D. A., & Lewandowsky, S. (2015). Memory without consolidation: Temporal distinctiveness explains retroactive interference. *Cognitive Science*, 39(7), 1570–1593. <https://doi.org/10.1111/cogs.12214>
- Erdfelder, E., Auer, T.-S., Hilbig, B. E., Aßfalg, A., Moshagen, M., & Nadarevic, L. (2009). Multinomial processing tree models: A review of the literature. *Zeitschrift für Psychologie/Journal of Psychology*, 217(3), 108–124. <https://doi.org/10.1027/0044-3409.217.3.108>
- Erdfelder, E., & Schnuerch, M. (2021). On the efficiency of the independent segments procedure: A direct comparison with sequential probability ratio tests. *Psychological Methods*, 26(4), 501–506. <https://doi.org/10.1037/met0000404>

- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. <https://doi.org/10.3758/BF03193146>
- Fillmore, M. T. (2001). Cognitive preoccupation with alcohol and binge drinking in college students: Alcohol-induced priming of the motivation to drink. *Psychology of Addictive Behaviors*, 15(4), 325–332. <https://doi.org/10.1037/0893-164X.15.4.325>
- Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4), 457–472. <https://doi.org/10.1214/ss/1177011136>
- Goodwin, D. W., Powell, B., Bremer, D., Hoine, H., & Stern, J. (1969). Alcohol and recall: State-dependent effects in man. *Science*, 163(3873), 1358–1360. <https://doi.org/10.1126/science.163.3873.1358>
- Hager, W., & Hasselhorn, M. (1994). *Handbuch deutschsprachiger Wortnormen [Handbook of German word norms]*. Hogrefe.
- Hardt, O., Nader, K., & Nadel, L. (2013). Decay happens: The role of active forgetting in memory. *Trends in Cognitive Sciences*, 17(3), 111–120. <https://doi.org/10.1016/j.tics.2013.01.001>
- Heck, D. W., Arnold, N. R., & Arnold, D. (2018). TreeBUGS: An R package for hierarchical multinomial-processing-tree modeling. *Behavior Research Methods*, 50(1), 264–284. <https://doi.org/10.3758/s13428-017-0869-7>
- Hu, X., & Batchelder, W. H. (1994). The statistical analysis of general processing tree models with the EM algorithm. *Psychometrika*, 59(1), 21–47. <https://doi.org/10.1007/BF02294263>
- Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430), 773–795. <https://doi.org/10.1080/01621459.1995.10476572>
- Keane, T. M., Lisman, S. A., & Kreutzer, J. (1980). Alcoholic beverages and their placebos: An empirical evaluation of expectancies. *Addictive Behaviors*, 5(4), 313–328. [https://doi.org/10.1016/0306-4603\(80\)90005-2](https://doi.org/10.1016/0306-4603(80)90005-2)
- Klauer, K. C. (2010). Hierarchical multinomial processing tree models: A latent-trait approach. *Psychometrika*, 75(1), 70–98. <https://doi.org/10.1007/s11336-009-9141-0>
- Knowles, S. K. Z., & Duka, T. (2004). Does alcohol affect memory for emotional and non-emotional experiences in different ways?. *Behavioural Pharmacology*, 15(2), 111–121. <https://doi.org/10.1097/00008877-200403000-00003>
- Kuhlmann, B. G., Brubaker, M. S., Pfeiffer, T., & Naveh-Benjamin, M. (2021). Longer resistance of associative versus item memory to interference-based forgetting, even in older adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 47(3), 422–438. <https://doi.org/10.1037/xlm0000963>
- Küpper-Tetzel, C. E., & Erdfelder, E. (2012). Encoding, maintenance, and retrieval processes in the lag effect: A multinomial processing tree analysis. *Memory*, 20(1), 37–47. <https://doi.org/10.1080/09658211.2011.631550>
- Lakens, D. (2019). The value of preregistration for psychological science: A conceptual analysis. *Japanese Psychological Review*, 62(3), 221–230. https://doi.org/10.24602/sjpr.62.3_221
- Lamberty, G. J., Beckwith, B. E., Petros, T. V., & Ross, A. R. (1990). Posttrial treatment with ethanol enhances recall of prose narratives. *Physiology & Behavior*, 48(5), 653–658. [https://doi.org/10.1016/0031-9384\(90\)90206-J](https://doi.org/10.1016/0031-9384(90)90206-J)
- Lohnas, L. J., Polyn, S. M., & Kahana, M. J. (2015). Expanding the scope of memory search: Modeling intralist and interlist effects in free recall. *Psychological Review*, 122(2), 337–363. <https://doi.org/10.1037/a0039036>
- Mann, K., Batra, A., Hoch, E., & die Leitliniengruppe. (2017). S3-Leitlinie "Screening, Diagnose und Behandlung alkoholbezogener Störungen": Kurzfassung [S3 guideline "screening, diagnosis, and treatment of alcohol-related disorders": Summary]. *SUCHT*, 63(1), 7–24. <https://doi.org/10.1024/0939-5911/a000464>
- Mann, R. E., Cho-Young, J., & Vogel-Sprott, M. (1984). Retrograde enhancement by alcohol of delayed free recall performance. *Pharmacology Biochemistry and Behavior*, 20(4), 639–642. [https://doi.org/10.1016/0091-3057\(84\)90317-4](https://doi.org/10.1016/0091-3057(84)90317-4)
- Manthey, J., Shield, K. D., Rylett, M., Hasan, O. S. M., Probst, C., & Rehm, J. (2019). Global alcohol exposure between 1990 and 2017 and forecasts until 2030: A modelling study. *The Lancet*, 393(10190), 2493–2502. [https://doi.org/10.1016/S0140-6736\(18\)32744-2](https://doi.org/10.1016/S0140-6736(18)32744-2)
- Mednick, S. C., Cai, D. J., Shuman, T., Anagnostaras, S., & Wixted, J. T. (2011). An opportunistic theory of cellular and systems consolidation. *Trends in Neurosciences*, 34(10), 504–514. <https://doi.org/10.1016/j.tins.2011.06.003>
- Mintzer, M. Z. (2007). The acute effects of alcohol on memory: A review of laboratory studies in healthy adults. *International Journal on Disability and Human Development*, 6(4), 397–403. <https://doi.org/10.1515/IJDHD.2007.6.4.397>
- Morey, R. D., & Rouder, J. N. (2018). *BayesFactor: Computation of Bayes factors for common designs*. R package version 0.9.12-4.2. <https://CRAN.R-project.org/package=BayesFactor>
- Moshagen, M. (2010). multiTree: A computer program for the analysis of multinomial processing tree models. *Behavior Research Methods*, 42(1), 42–54. <https://doi.org/10.3758/BRM.42.1.42>
- Mueller, C. W., Lisman, S. A., & Spear, N. E. (1983). Alcohol enhancement of human memory: Tests of consolidation and interference hypotheses. *Psychopharmacology*, 80(3), 226–230. <https://doi.org/10.1007/BF00436158>
- Parker, E. S., Birnbaum, I. M., Weingartner, H., Hartley, J. T., Stillman, R. C., & Wyatt, R. J. (1980). Retrograde enhancement of human memory with alcohol. *Psychopharmacology*, 69(2), 219–222. <https://doi.org/10.1007/BF00427653>
- Parker, E. S., Morihisa, J. M., Wyatt, R. J., Schwartz, B. L., Weingartner, H., & Stillman, R. C. (1981). The alcohol facilitation effect on memory: A dose-response study. *Psychopharmacology*, 74(1), 88–92. <https://doi.org/10.1007/BF00431763>
- Quevedo Pütter, J., Erdfelder, E., & Kieslich, P. J. (2020). *Registered report protocol preregistration: Does alcohol consumption after learning really improve memory? And if so, why?* <https://osf.io/dk8fj>
- Rasch, B., & Born, J. (2013). About sleep's role in memory. *Physiological Reviews*, 93, 681–766. <https://doi.org/10.1152/physrev.00032.2012>
- Roehrs, T., & Roth, T. (2001). Sleep, sleepiness, and alcohol use. *Alcohol Research & Health*, 25(2), 101–109. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6707127/>
- Rouder, J. N., & Batchelder, W. H. (1998). Multinomial models for measuring storage and retrieval processes in paired associate learning. In C. E. Dowling, F. S. Roberts, & P. Theuns (Eds.), *Recent progress in mathematical psychology: Psychophysics, knowledge, representation, cognition, and measurement* (pp. 195–225). Lawrence Erlbaum Associates Publishers.
- Saunders, J. B., Aasland, O. G., Babor, T. F., de la Fuente, J. R., & Grant, M. (1993). Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO collaborative project on early detection of persons with harmful alcohol consumption-II. *Addiction*, 88(6), 791–804. <https://doi.org/10.1111/j.1360-0443.1993.tb02093.x>
- Schmidt, O., Erdfelder, E., & Heck, D. W. (2023). How to develop, test, and extend multinomial processing tree models: A tutorial. *Psychological Methods*. Advance online publication. <https://doi.org/10.1037/met0000561>
- Schnuerch, M., & Erdfelder, E. (2020). Controlling decision errors with minimal costs: The sequential probability ratio t test. *Psychological Methods*, 25(2), 206–226. <https://doi.org/10.1037/met0000234>
- Schnuerch, M., Heck, D. W., & Erdfelder, E. (2022). Waldian t tests: Sequential Bayesian t tests with controlled error probabilities. *Psychological Methods*. Advance online publication. <https://doi.org/10.1037/met0000492>

- Steinhilber, M., Schnuerch, M., & Schubert, A.-L. (2021). *sprtt: Sequential Probability Ratio Tests: Using t-Statistic*. R package version 0.1.0. <https://CRAN.R-project.org/package=sprtt>
- Thierauf, A., Kempf, J., Eschbach, J., Auwärter, V., Weinmann, W., & Gnann, H. (2013). A case of a distinct difference between the measured blood ethanol concentration and the concentration estimated by Widmark's equation. *Medicine, Science and the Law*, 53(2), 96–99. <https://doi.org/10.1258/msl.2012.012038>
- Tucker, M. A., Hirota, Y., Wamsley, E. J., Lau, H., Chaklader, A., & Fishbein, W. (2006). A daytime nap containing solely non-REM sleep enhances declarative but not procedural memory. *Neurobiology of Learning and Memory*, 86(2), 241–247. <https://doi.org/10.1016/j.nlm.2006.03.005>
- Tulving, E., & Psotka, J. (1971). Retroactive inhibition in free recall: Inaccessibility of information available in the memory store. *Journal of Experimental Psychology*, 87(1), 1–8. <https://doi.org/10.1037/h0030185>
- Tyson, P. D., & Schirmuly, M. (1994). Memory enhancement after drinking ethanol: Consolidation, interference, or response bias? *Physiology & Behavior*, 56(5), 933–937. [https://doi.org/10.1016/0031-9384\(94\)90326-3](https://doi.org/10.1016/0031-9384(94)90326-3)
- Weafer, J., Gallo, D. A., & De Wit, H. (2016a). Acute effects of alcohol on encoding and consolidation of memory for emotional stimuli. *Journal of Studies on Alcohol and Drugs*, 77(1), 86–94. <https://doi.org/10.15288/jsad.2016.77.86>
- Weafer, J., Gallo, D. A., & De Wit, H. (2016b). Effect of alcohol on encoding and consolidation of memory for alcohol-related images. *Alcoholism: Clinical and Experimental Research*, 40(7), 1540–1547. <https://doi.org/10.1111/acer.13103>
- Wechsler, H., Dowdall, G. W., Davenport, A., & Rimm, E. B. (1995). A gender-specific measure of binge drinking among college students. *American Journal of Public Health*, 85(7), 982–985. <https://doi.org/10.2105/AJPH.85.7.982>
- White, N. M. (1996). Addictive drugs as reinforcers: Multiple partial actions on memory systems. *Addiction*, 91(7), 921–950. <https://doi.org/10.1046/j.1360-0443.1996.9179212.x>
- Wixted, J. T. (2004). The psychology and neuroscience of forgetting. *Annual Review of Psychology*, 55, 235–269. <https://doi.org/10.1146/annurev.psych.55.090902.141555>
- Wixted, J. T. (2010). The role of retroactive interference and consolidation in everyday forgetting. In S. Della Sala (Ed.), *Forgetting* (pp. 285–312). Psychology Press.

History

Received December 31, 2019
 Revision received September 29, 2022
 Accepted December 12, 2022
 Published online February 21, 2023

Acknowledgments

We would like to thank Pascal J. Kieslich for his help with preparing the study, and Lena Bizer, Annalena Loose, Linus Quevedo Pütter, Pia Quevedo Pütter, Christina Sarafoglou, and Julian Ziegler for their help with data collection.

Publication Ethics

The research protocol was approved by the ethics committee of the University of Mannheim.



Open Data

The Stage-1 manuscript of this registered report (<https://doi.org/10.17605/OSF.IO/DK8FJ>; Quevedo Pütter et al., 2020), the materials (<https://doi.org/10.17605/OSF.IO/8E9PW>), and the data (<https://doi.org/10.17605/OSF.IO/2K4J7>) are publicly available on the Open Science Framework (OSF). Additional information needed to reproduce all reported results is provided in the electronic supplemental material (ESM).

Funding

This research was funded by the Deutsche Forschungsgemeinschaft (DFG), Grant 2277, Research Training Group “Statistical Modeling in Psychology” (SMiP). Open access publication enabled by University of Mannheim.

ORCID

J. Quevedo Pütter
 <https://orcid.org/0000-0002-7340-9937>
 E. Erdfelder
 <https://orcid.org/0000-0003-1032-3981>

Edgar Erdfelder

Cognition and Individual Differences Lab
 University of Mannheim
 68131 Mannheim
 Germany
erdfelder@uni-mannheim.de