

## RESEARCH ARTICLE

## Examining the effects of time of day and sleep on generalization

Marlie C. Tandoc<sup>1\*</sup>, Mollie Bayda<sup>2,3</sup>, Craig Poskanzer<sup>2,4</sup>, Eileen Cho<sup>2</sup>, Roy Cox<sup>2,5</sup>, Robert Stickgold<sup>2</sup>, Anna C. Schapiro<sup>1,2\*</sup>

**1** Department of Psychology, University of Pennsylvania, Philadelphia, Pennsylvania, United States of America, **2** Department of Psychiatry, Beth Israel Deaconess Medical Center / Harvard Medical School, Boston, Massachusetts, United States of America, **3** Department of Psychology, University of California-Los Angeles, Los Angeles, California, United States of America, **4** Department of Psychology and Neuroscience, Boston College, Chestnut Hill, Massachusetts, United States of America, **5** Department of Sleep and Cognition, Netherlands Institute for Neuroscience, Amsterdam, The Netherlands

\* [tandoc@sas.upenn.edu](mailto:tandoc@sas.upenn.edu) (MCT); [aschapiro@sas.upenn.edu](mailto:aschapiro@sas.upenn.edu) (ACS)



## OPEN ACCESS

**Citation:** Tandoc MC, Bayda M, Poskanzer C, Cho E, Cox R, Stickgold R, et al. (2021) Examining the effects of time of day and sleep on generalization. *PLoS ONE* 16(8): e0255423. <https://doi.org/10.1371/journal.pone.0255423>

**Editor:** Bradley R. King, University of Utah, UNITED STATES

**Received:** March 28, 2021

**Accepted:** July 16, 2021

**Published:** August 2, 2021

**Peer Review History:** PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0255423>

**Copyright:** © 2021 Tandoc et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the manuscript and its [Supporting Information](#) files.

**Funding:** This work was supported by: NIH NINDS F32-NS093901 (ACS); MH048832 (RS) The

## Abstract

Extracting shared structure across our experiences allows us to generalize our knowledge to novel contexts. How do different brain states influence this ability to generalize? Using a novel category learning paradigm, we assess the effect of both sleep and time of day on generalization that depends on the flexible integration of recent information. Counter to our expectations, we found no evidence that this form of generalization is better after a night of sleep relative to a day awake. Instead, we observed an effect of time of day, with better generalization in the morning than the evening. This effect also manifested as increased false memory for generalized information. In a nap experiment, we found that generalization did not benefit from having slept recently, suggesting a role for time of day apart from sleep. In follow-up experiments, we were unable to replicate the time of day effect for reasons that may relate to changes in category structure and task engagement. Despite this lack of consistency, we found a morning benefit for generalization when analyzing all the data from experiments with matched protocols ( $n = 136$ ). We suggest that a state of lowered inhibition in the morning may facilitate spreading activation between otherwise separate memories, promoting this form of generalization.

## Introduction

No two of our experiences are exactly alike. Every decision we make thus involves some degree of generalization. Tracking the shared properties across our experiences allows us to flexibly apply our prior knowledge to novel situations. For example, learning that oil and butter can provide a similar use when frying eggs may allow you to infer that they could support similar functions elsewhere, such as when baking a cake or removing a stuck ring. We will examine how this form of generalization, involving recombinations of previously separate memories, varies as a function of different brain states: offline during sleep, and online at different times of the day.

fundings had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

Sleep plays an important role not only in the preservation of our memories, but also in their reorganization [1, 2]. For example, sleep benefits knowledge of shared features [3, 4] and statistical regularities [5] across recently learned items and facilitates the integration of memories into existing knowledge [6, 7]. Sleep may promote the formation of neocortical representations that reflect that shared structure across experiences, which should afford generalization [8].

Despite these findings that sleep can enhance knowledge of structured information, evidence for the benefits of sleep on generalization, the act of applying that knowledge to novel contexts, is more mixed [3, 4, 9–16]. Discovering the kinds of generalization that are enhanced by sleep, and developing behavioral measures that are sensitive to this process, is critical for understanding how sleep-related memory reorganization affects the ability to generalize. We test the idea that generalization across concepts previously seen as unrelated might particularly benefit from the increased representational overlap that emerges during sleep, as they otherwise would be represented distinctly. To this end, we employ a novel category learning paradigm that provides minimal opportunity to extract shared structure during initial learning.

Some forms of generalization, including the integration of information across otherwise separate memories can occur online during a learning experience [17, 18]. The brain states that fluctuate with circadian rhythms, the physiological changes that cycle across the day, may influence this in-the-moment generalization. Within the sleep literature, time of day effects have arisen unexpectedly in tasks that involve generalization. In these studies, categorization, which often involves making decisions about unseen exemplars, was found to be better in the morning [19, 20]. In addition, the tendency to generalize fear extinction responses across contexts is stronger in the morning [15, 21]. The ability to generalize during speech comprehension also appears enhanced in the morning [22]. Together, these studies already provide some evidence that generalization may vary across the day, with superior performance in the morning.

Although these effects were mostly considered secondary curious results, there is a rich literature demonstrating time of day effects on cognition. Several cognitive processes vary across the circadian cycle in both animals [23] and humans [24], including attention [25–27], executive functions [28–39], and various forms of long-term memory [40–53], including semantic memory [54–58].

What leads to time of day effects on cognition? A long-standing framework posits that inhibitory cognitive processes are strongest at peak times of circadian arousal and weakest at off-peak times [32, 34, 39, 59]. These optimal times of day are determined by an individual's chronotype (morning or evening preference) [60], which is typically the evening for young adults [39, 45, 59]). Evidence that off-peak circadian states may be associated with low inhibition include impaired performance on tasks with high cognitive control demands [37, 50, 61] and greater interference between memories [35, 62]. Although interference between memories can be deleterious for their veridical preservation, this lower inhibitory state may be beneficial for connecting memories to uncover shared structure. Indeed, aspects of cognition that rely on noticing connections between otherwise distinct events, such as insight problem-solving in the Remote Associates Test, which involves seeking a word that connects three seemingly unrelated words, show improvements at off-peak times of day [63, 64]. Consistent with theories arguing that reduced cognitive control can be adaptive [65, 66], we suggest that generalization may be another instance where low inhibition is an asset [67]. Specifically, lower inhibition in the morning might facilitate the spreading of activation across different memories, making it easier to see novel connections.

To assess impacts of sleep and time of day on generalization, we developed a variant of a category learning paradigm [4] in which participants learned about three categories of novel satellites. After these satellites were well-learned, participants encountered two *bridge* satellites,

which combined features from two of the initially learned categories. We assessed how encountering examples of satellites that combine features across previously separate categories facilitates the ability to generalize to novel satellites combining features from the same categories. Given that generalization can coincide with false memory formation [68, 69, cf. 70, 71], we also examined false memories for these novel bridge satellites.

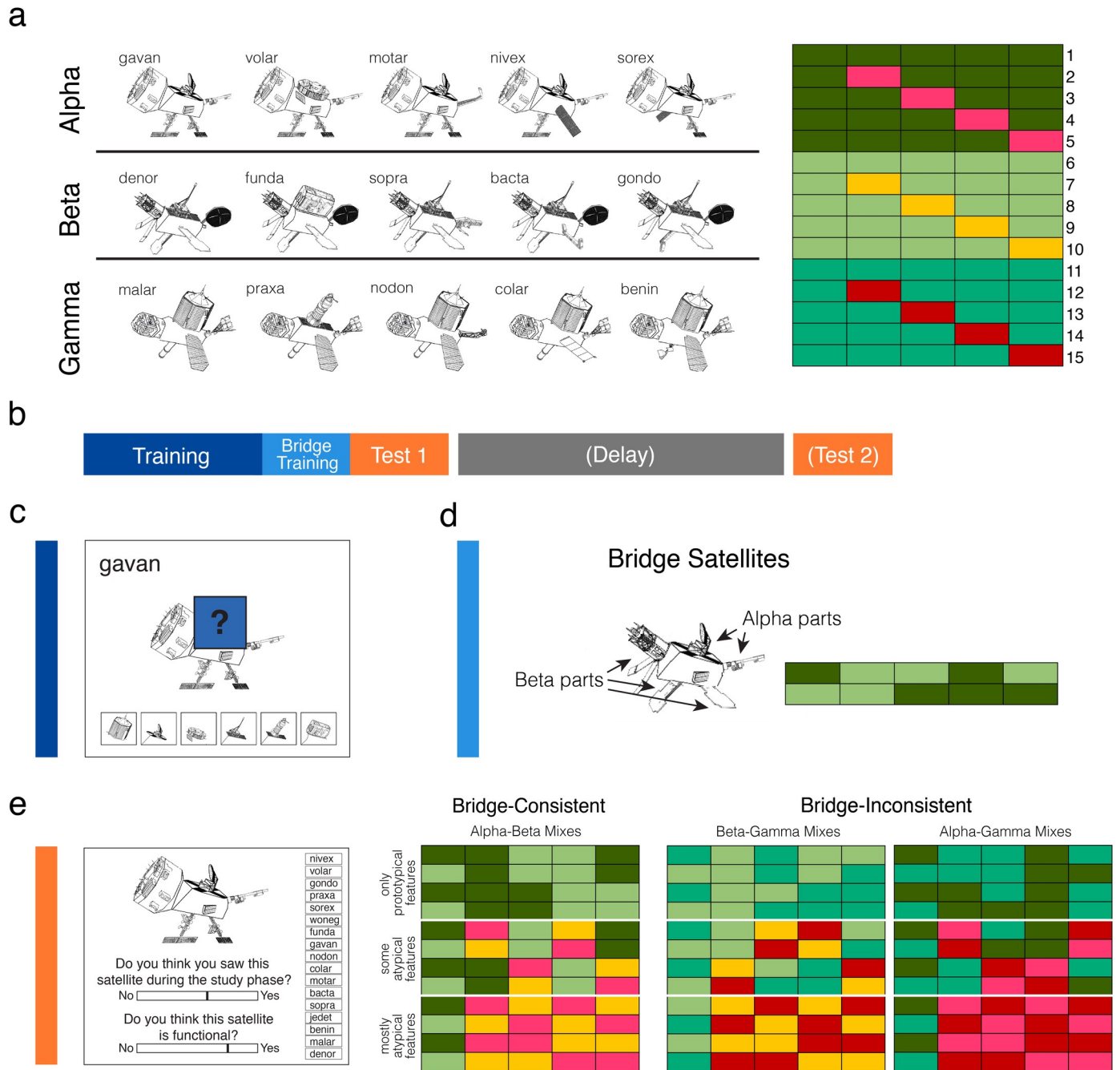
Our first three experiments were designed as sleep studies, involving two sessions spaced 12 or 24 hours apart (Experiments 1–3). Although we hypothesized that sleep may benefit this form of generalization, we were unable to detect any evidence of such an effect in these experiments. Instead, we observed a time of day effect, with better generalization in the morning. A nap study suggested that this effect is likely due to the particular time of day, as opposed to having slept recently (Experiment 4). In the next two experiments (Experiments 5 and 6), we pursued this effect using a simplified paradigm and administered additional tasks, the RAT and distractor priming, known to vary with time of day. Finally, we attempted to replicate the time of day effect in our original generalization paradigm, with the RAT and distractor priming task additionally administered at the end (Experiment 7). In these last three experiments, we did not replicate the time of day effect, possibly due to changes to the protocol or to task engagement.

## Methods

### Experiment 1

**Participants.** 33 participants from the Boston area (24 females, age = 21.91, range = 18–34) participated in exchange for monetary compensation. They were randomly assigned to a Morning group or Evening group. For consistency across experiments, conditions are named according to the time of the first session. Data from 3 participants were excluded because in the first test phase they were unable to discriminate novel from trained satellites (mean recognition for novel satellites > mean recognition for trained satellites) or they were unable to reliably recognize trained satellites (mean recognition for trained satellites below baseline), leaving 13 participants in the Morning group and 17 in the Evening group. The same exclusion criteria were implemented in all experiments. Across experiments, differences in sample size between conditions and in some cases limited sample sizes arose due to institution scheduling constraints, such as the availability of required nurse support, and participant cancellations. Participants completed a pre-enrollment screening survey and were only enrolled if they reported: keeping a normal sleep schedule (going to bed before 2am and sleeping at least 6 hours per night); having no psychiatric, neurological, or sleep-related disorders; not taking medications known to interfere with sleep; drinking less than four servings of caffeine per day and less than 10 alcoholic beverages per week; and having normal or corrected-to-normal vision and hearing. The study was approved by the Institutional Review Board of Beth Israel Deaconess Medical Center. All subjects provided written informed consent, and the experiment was carried out in accordance with relevant guidelines and regulations.

**Stimuli.** The main training stimuli were 15 novel “satellite” objects drawn from three categories, “Alpha,” “Beta,” and “Gamma” (Fig 1A). Each satellite had five parts: head, back, front leg, back leg, and tail. There were five members of each category, with satellites in the same category sharing most of the five parts. Each category contained one prototype, and each of the other satellites in the category had one part deviating from that prototype. Satellites from different categories did not share any parts. Each satellite had shared features: the “class” name and the parts shared among members of the category; it also had unique features: a unique “code” name and the part unique to that satellite (except for the prototype, which had no



**Fig 1. Stimuli and task design.** (a) Examples of the 15 training exemplars that made up the three categories (left). Category structure of the 15 training exemplars (right) depicting the features (columns) that made up each exemplar (rows). Rows 1–5 are Alpha exemplars, 6–10 Beta, and 11–15 Gamma. Shades of green depict category prototypical features and non-green colors indicate category atypical features. (b) General structure of experiments. (c) Example of a training trial, where participants inferred missing features. (d) Example of a bridge satellite combining prototypical Alpha and Beta features and the feature structure of the two bridge satellites shown during bridge training. Participants learned these satellites in the same way as in the training phase. (e) Schematic of test trial (left) which involves making slider judgments about whether that particular satellite was seen during training (recognition) and if the satellite was functional (functional judgment). Participants might then guess the exemplar code name by clicking on the name (code name memory). Feature structure is depicted (right) for the test items which combined features of the initially learned categories. These satellites could combine features from the same two categories shown during bridge training (bridge-consistent) or from other two pairings of categories (bridge-inconsistent). Test items also could vary in the typicality of their features (only prototypical, some atypical, mostly atypical).

<https://doi.org/10.1371/journal.pone.0255423.g001>

unique part). Satellites were constructed randomly for each participant, constrained by this category structure.

Participants also learned about two “bridge” satellites, which combined prototypical features from two of the categories (Fig 1D). One of the bridge items contained three features from one category and two from the other, and vice versa for the second bridge item.

At test, participants encountered the 17 satellites described above, in addition to 36 novel satellites with the structures shown in Fig 1E. These satellites combined features from each pair of categories with varying use of atypical features: four of the mixes involved only prototypical features from the two categories, four involved some atypical features, and four were mostly atypical features. In total, this produced 12 bridge-consistent and 24 bridge-inconsistent satellites as test items. Test items were carefully designed such that the satellites from each of the three types of mixes had equal amounts of feature overlap with the bridge items.

**Procedure.** Participants were asked not to consume alcohol or recreational drugs in the 24 hours prior to the study, and to abstain from caffeine on the day of the study. They were also asked to maintain a regular sleep schedule in the three days prior to the study, with no bedtimes earlier than 10pm or later than 2am. In the Evening condition, participants came to the lab at 8pm for the training phase. Before beginning, they filled out surveys, including a retrospective three-day sleep diary, the Epworth Sleepiness Scale [72], and the Stanford Sleepiness Scale (SSS; [73]). When finished with the training phase, around 9pm, they completed the immediate test. They slept in the lab with polysomnographic (PSG) recording and filled out the SSS again and were tested again at 9am. The PSG data were not analyzed due to lack of behavioral evidence for sleep effects. In the Morning condition, participants came to the lab at 9am for the training phase. At around 10am they completed the immediate test. They left the lab to participate in normal daily activities except they were instructed not to nap, and they came back to the lab at 9:30pm for a delayed test. The general experimental procedure is depicted in Fig 1B.

During the training phase, participants were first taught about the parts belonging to the 15 satellite objects. Participants were told that “these 15 satellites are functional, which means the parts can work properly together.” They were also told that the satellites belong to three classes. Participants were first shown each of the 15 satellites one at a time for five seconds each. The display included the class name, code name, and full satellite image.

The experiment then moved on to the main portion of the training phase, a property inference task. On each trial, participants were shown a satellite with its code name displayed as well as four out of five of its visual features (Fig 1C). A blue box with a question mark occluded one of the visual features and participants were given six options at the bottom as to what the occluded feature might be. These options consisted of all possible versions of that feature across the 15 satellites, including the three prototypical and three atypical variants. Participants used the mouse to click on one of the feature options and received immediate feedback about whether their choice was correct, and the occluding box was removed to reveal the true feature. If the choice was incorrect, the trial would start over with shuffled locations of the feature options, and the participant would continue repeating the trial until they chose the correct feature.

Once the correct feature was chosen, the three possible class names, “Alpha,” “Beta,” and “Gamma,” were displayed, and the participant was prompted to choose the correct name. Participants continued to select class names, with feedback as to whether their response was correct, until clicking the correct name.

The training phase proceeded in blocks, with 45 trials per block. Each block contained three trials for each of the fifteen satellites, two testing each satellite’s unique feature and one testing a shared feature. Unique features were queried twice as often as shared, since shared

features are reinforced across members of the category and are thus easier to learn. At the end of each block, participants were shown their accuracy for the visual feature guessing and for the class name guessing. They were told that they needed to reach 66% correct across trials on the visual parts and 90% correct on the class names. Once they reached both criteria or one hour had passed, they moved on to the bridge training phase.

In the bridge training phase, participants learned about the two bridge items, which combined features from two of the categories (Fig 1D). Participants were told: “You have now studied 15 satellites. In this next part of the study phase, we are interested in how difficult it will be to now learn about two new satellites. Each of these new satellites has combinations of parts that are functional.” This was simply a cover story, as we did not want participants to be clued in to the structure of the bridge items. Participants were then shown the code names and images of the two new satellites (no class names). They then did the same property inference task as before except that they did not guess class names. There were blocks of ten trials to query each of the five features from the two satellites, with trials alternating between the two satellites. They repeated the block until identifying nine out of ten features correctly.

Participants then completed the test phase, which had 53 trials. These trials consisted of the 15 trained satellites and 2 trained bridge items, as well as 36 novel satellites. On each trial, participants saw a satellite with one slider bar below asking if they had seen that satellite during the study phase (Fig 1E). The ends of the scales were labeled “Definitely not” and “Definitely yes” and the middle was labeled “Not sure.” They clicked a location on the slider and then either a “submit” button to register the response or a “reset” button, allowing them to change their response before submitting. As long as the slider response was not placed within the bottom 5% of the scale (indicating that they definitely did not recognize the satellite), 17 options for code names appeared on the right side of the screen, and participants clicked to choose an option. They could then submit or reset. Subjects next saw a second slider bar appear below the first, prompting a response on whether the satellite was functional. The following note was always displayed next to this slider: “Please try to make use of the entire slider. (All satellites that you have studied are functional. Many of the satellites that you have not studied are also functional.)” Slider values were recorded in the range -1 (Definitely not) to 1 (Definitely yes).

For the delayed test approximately 12 hours later, participants completed the same test phase with the trials in a different random order. The experiments were implemented in MATLAB with Psychophysics Toolbox [74].

**Generalization index calculation.** For each participant, a generalization index was calculated as a difference score (mean slider ratings for novel bridge-consistent Satellites–mean slider ratings for novel bridge-inconsistent satellites) separately for their functional slider ratings and recognition slider ratings. For functional judgment generalization, a positive index indicates that a participant endorses novel bridge-consistent satellites (satellites that share features from the two categories shown to produce a functional satellite during the bridge phase) as more functional relative to novel bridge-inconsistent satellites (satellites that combine features from categories never shown to produce a functional satellite). For recognition generalization, a positive index indicates that a participant false alarms to novel bridge-consistent satellites more than to novel bridge-inconsistent satellites.

## Experiment 2

31 participants were recruited (18 females, age = 22.23, range = 18–35) and 4 were excluded according to the criteria specified above. This experiment was the same as Experiment 1 except that the number of blocks of exposure for the bridge items was not determined by performance but instead fixed separately for each group in an attempt to match performance between

groups. The Morning group ( $n = 14$ ) only saw one block of bridge items and the Evening group ( $n = 13$ ) saw six blocks. Given the effect size for functional judgment generalization observed in Experiment 1 (Cohen's  $d = 1.20$ ), this sample provides over 90% power to detect the same effect. In Experiment 1 the Morning group took  $2.46 \pm 0.27$  blocks to learn and the Evening group took  $2.59 \pm 0.32$  blocks. Relative to Experiment 1, the Experiment 2 Morning group thus received less bridge training and the Experiment 2 Evening group received more bridge training.

### Experiment 3

This experiment was the same as Experiment 1 except that the delayed test took place 24 hours later instead of approximately 12 hours later. Participants did not sleep in the lab. From this experiment on, we also collected the Morningness-Eveningness Questionnaire [60] at the end of the protocol for each participant (MEQ). 29 participants were recruited (15 females, age = 23.03, range = 18–32). Two participants failed to meet the inclusion criteria, leaving 14 participants in the Morning group and 13 participants in the Evening group.

### Experiment 4

This experiment was the same as Experiment 1 except that all participants began training at 4pm, and there was no delayed test. Participants were randomly assigned to either take a nap in the lab with PSG recording prior to the test (Nap group) or to come into the lab without having taken a nap (No Nap group). 41 participants were recruited (23 females, age = 23.39, range 18–33) and 4 were excluded, leaving 20 participants in the Nap group and 17 in the No Nap group. One participant's total sleep time was manually recorded due to a technical error.

### Experiment 5

This experiment was the same as Experiment 1 except that there was again no delayed test, there were two additional new tasks administered before the satellite task, and the satellite task was simplified to allow time for the additional tasks. Specifically, there were now three satellites instead of five in each of the three categories. Using the numbering shown in Fig 1A, items 3, 5, 7, 9, 12, and 15 were removed. There were now 27 trials per block during the initial training phase. The test was also shortened: All generalization test trials with atypical features were removed, resulting in a total of 21 trials. 58 participants were recruited (43 females, age = 21.68, range = 18–30) and 15 were excluded (1 due to procedural difficulties), leaving 25 in the Morning group and 18 in the Evening group.

The two new tasks were the Remote Associates Task (RAT; [75]) and a distractor priming task [37, 76]. Participants first completed the RAT, in which on every trial they saw three words and had to find a fourth that connects them by forming a two-word phrase or a compound word. Participants were given the example *Eight / Skate / Stick*, with the solution *Figure*. They completed two practice trials followed by 20 test trials. They were given a maximum of 30 seconds to type in a solution word on each trial. If they typed the correct word with errors, partial credit was assigned using Levenshtein distance, a string comparison metric that calculates the minimum number of character changes needed to transition from one word (the typed response) to another (the solution word). Only 2.5% of participant responses were assigned partial credit using this method. Manually scoring these partial credits as either correct or incorrect for semantic accuracy (ignoring spelling and typographical errors) had no material impact on the results.

The distractor priming task had two parts. The first part was a 1-back task, in which participants were presented with a series of red line drawings of familiar objects and animals.

Superimposed on the pictures were words and nonwords in black font. Each stimulus pair was displayed for 1000 ms, followed by a blank screen for 500 ms. Participants were told to ignore the words and to press one key whenever the current drawing was identical to the previous one, and a different key otherwise. After a practice phase, there were 69 trials of the task, with 15 target words, 15 fillers, and 39 nonwords. The identity of the target words and fillers were counterbalanced across subjects. About one in every seven trials had a repeating image.

The second part of the distractor priming task was presented to participants as a completely unrelated task. In this part, participants were told they would see a “mystery” word with some blank letters and to quickly type in the first word that comes to mind to complete the missing letters: “For example, if you see T\_BL\_, you might type TABLE and then press enter.” Participants were given a maximum of 5 s to respond on each trial. There were 15 fragments that could be completed by a target word from the prior task, 15 fragments of words from the other counterbalanced group’s targets, and 15 easily-solvable filler fragments. Our implementation of the distractor priming task differed from prior implementations in that there was no delay between the two parts of the task, and participants typed their fragment completion responses instead of speaking them outloud.

To generate a priming score, we compared the typed response to the target word on each trial. If the participant typed the correct target word, the response received a score of 1. If they typed a different word or did not type a word, the response received a score of 0. Otherwise, we used Levenshtein distance to provide partial credit. Priming was then calculated as a difference score for each participant between their mean score to targets and mean score to novel items (which served as targets for the counterbalanced group). As prior studies have calculated this measure slightly differently, contrasting mean scores to targets with means scores on these same words from the counterbalanced group, we also computed the priming score in this manner and found that it did not change the results.

## Experiment 6

This experiment was the same as Experiment 5 except that the RAT and priming test were administered after (rather than before) the satellite task. 53 participants were recruited (42 females, age = 21.69, range = 18–34) and 4 were excluded, leaving 26 participants in the Morning group and 23 in the Evening group. One participant did not complete surveys, but was still included in analyses.

## Experiment 7

This experiment was the same as Experiment 6 except that the original stimuli and test set, with 5 satellites per category and 53 test items, were used for the satellite task. 86 participants were recruited (60 females, age = 21.74, range = 18–36) and 7 were excluded, leaving 41 participants in the Morning group and 38 participants in the Evening group. One participant did not complete surveys and another participant was excluded from the priming test due to procedural difficulties.

Design differences for each experiment, as well as the observed time of day effects on generalization, are summarized in [Table 1](#).

## Results

### Experiment 1

**Training performance and surveys.** Training phase results, including the mean proportion correct on the last block of initial training for feature and class accuracy, the number of



Table 1. Experiment design and time of day effects on generalization.

	Experiment Design								Time of day effect on generalization	
	Number of Sessions	Delay Between Sessions	Session 1 Start Time	Session 2 Start Time	Bridge Exposure	Exemplars Per Category	Number of Test Trials	Order of Tasks	Functional judgment <i>p</i>	Recognition <i>p</i>
Exp. 1	2	12 hours	Morning: 9AM Evening: 8PM	Morning: 9:30PM Evening: 9AM	To criterion	5	53		.003**	.06§
Exp. 2	2	12 hours	Morning: 9AM Evening: 8PM	Morning: 9:30PM Evening: 9AM	Morning: 1 block Evening: 6 blocks	5	53		.44	.52
Exp. 3	2	24 hours	Morning: 9AM Evening: 8PM	Morning: 9AM Evening: 8PM	To criterion	5	53		.0003***	.01*
Exp. 4	1		4PM		To criterion	5	53			
Exp. 5	1		Morning: 9AM Evening: 8PM		To criterion	3	21	RAT, Distractor Priming, Generalization	.08§	.18
Exp. 6	1		Morning: 9AM Evening: 8PM		To criterion	3	21	Generalization, RAT, Distractor Priming	.47	.34
Exp. 7	1		Morning: 9AM Evening: 8PM		To criterion	5	53	Generalization, RAT, Distractor Priming	.56	.93
Exp. 1, 3, & 7 Combined			Morning: 9AM Evening: 8PM		To criterion	5	53		.0007***	.04*

Number of Sessions = number of experimental sessions; two session experiments had a second test administered after a delay. Delay Between Sessions = approximate time between Session 1 and Session 2. Session 1 Start Time = approximate start time of Session 1 for each group. Session 2 Start Time = approximate start time of Session 2 (referred to in text as Test 2) for each group. Bridge Exposure = number of blocks (10 trials per block) that participants encountered bridge satellites during the training phase. Criterion was reached when a participant achieved nine out of ten trials in a block correct. In Experiment 2 the number of blocks was different between groups, reflecting our aim to match initial generalization performance in this experiment. Exemplars Per Category = the number of unique satellites in each of the three categories encountered during the training phase. Number of Test Trials = total number of trials during the test phase in one session. Order of Tasks = the order of the tasks that participants completed in experiments that administered additional measures (RAT, Distractor Priming). Generalization refers to the full generalization paradigm (i.e. both training and test). Functional judgment *p* = Difference between morning and evening group for functional judgment generalization. Recognition *p* = Difference between morning and evening group for recognition generalization. \*\*\**p* < .001, \*\**p* < .01, \**p* < .05, §*p* < .1

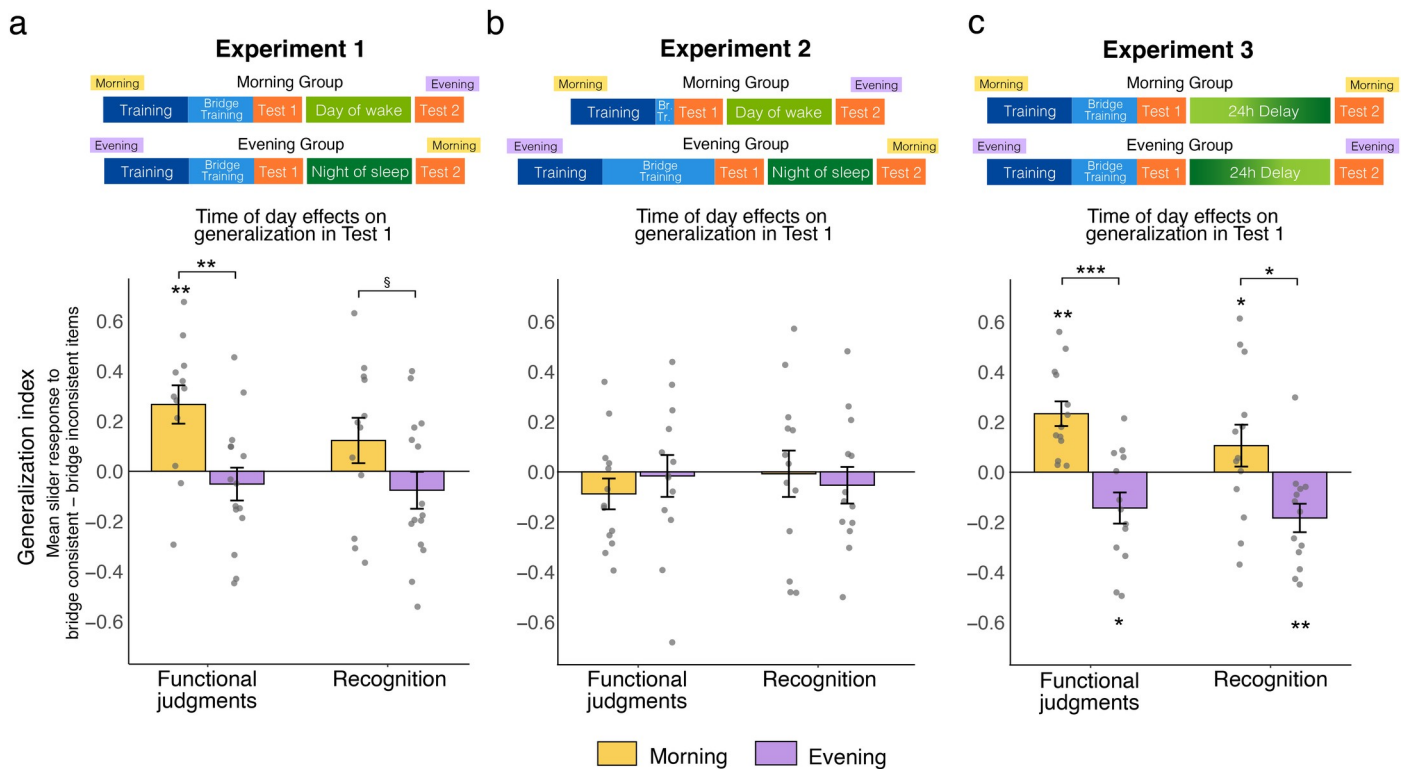
<https://doi.org/10.1371/journal.pone.0255423.t001>

blocks required to learn the initial group of satellites as well as the bridge satellites are reported in S1 Table. We did not find any differences between groups. Refer to S2 Table for survey measures for all experiments, including Stanford Sleepiness Scale reports.

**Basic measures of performance in Test 1.** We first assessed performance on functional and recognition judgments for trained satellites (the 17 satellites from the initial and bridge phases). While there was no difference between groups on recognition judgments for trained satellites ( $t[28] = 0.36, p = .72$ ), the Morning group rated trained satellites as more functional than the Evening group ( $t[28] = 2.79, p = .009$ ; Morning:  $0.72 \pm 0.05$ ; Evening:  $0.55 \pm 0.04$ ). The groups did not differ in their mean overall functional ( $t[28] = 0.74, p = .47$ ) or recognition ( $t$

[28] = -1.58,  $p = .13$ ) slider responses. We found no difference between groups in the proportion of correct responses for code name memory ( $t[28] = 0.07, p = 0.94$ ). Overall, basic measures of performance in Test 1 did not differ between groups, except that the Morning group rated trained satellites as more functional.

**Time of day effects on generalization in Test 1.** Before assessing the role of sleep on generalization, we sought to rule out any differences between groups on initial generalization performance in Test 1. Strikingly, functional judgment generalization was significantly different between the groups tested in the morning and evening ( $t[28] = 3.28, p = .003$ ) whereby the Morning group was better able to attribute the functional property to novel bridge-consistent satellites than the Evening group (Fig 2A). The Morning group also exhibited marginally higher recognition generalization than the Evening group ( $t[28] = 1.92, p = .06$ ), suggesting that participants tested in the morning had more false memories for novel bridge-consistent satellites than those tested in the evening. Furthermore, only the Morning group exhibited positive generalization for functional judgments (Morning:  $t[12] = 3.47, p = .005$ , Evening:  $t[16] = -0.88, p = 0.39$ ), indicating that only participants tested in the morning reliably rated bridge-consistent satellites as more functional than bridge-inconsistent ones. Neither group showed reliably positive generalization for the recognition judgment (Morning:  $t[12] = 1.55, p = .15$ ; Evening:  $t[16] = -1.09, p = .29$ ). There was a positive correlation between functional judgment generalization and recognition generalization ( $r = 0.73, p = .00005$ ), suggesting that if a participant was likely to endorse novel bridge-consistent satellites as functional, they were also more likely to falsely endorse novel bridge-consistent satellites as having been seen during



**Fig 2. Time of day effects on generalization in Test 1.** (a) Experiment 1. (b) Experiment 2, where bridge training was fixed to match initial generalization between groups: the morning group received only one block of bridge-satellite training whereas the evening group received six blocks. This is in contrast to Experiment 1 and 3 where participants trained to criterion. (c) Experiment 3. A schematic of the timeline (not to scale) for each experiment is shown above. Gray dots correspond to individual participants and error bars denote  $\pm$  one SEM. \*\*\* $p < .001$ , \*\* $p < .01$ , \* $p < .05$ , \$ $p < .1$ .

<https://doi.org/10.1371/journal.pone.0255423.g002>

training. Time since awakening from sleep did not correlate with either measure of generalization for the Morning or Evening group ( $p$ 's > .57).

The novel bridge satellite test items varied in the typicality of their features (referred to as feature type). Regressions with time of day and feature type (ordered as prototypical, some atypical, mostly atypical) as predictors revealed a significant main effect of time of day on functional judgment generalization ( $B = 0.15$ ,  $t = 3.51$ ,  $p = .0007$ ) and on recognition memory generalization ( $B = 0.10$ ,  $t = 2.15$ ,  $p = .03$ ). There was no main effect of feature type, nor was there any interaction between time of day and feature type on either measure of generalization ( $p$ 's > .42). These results indicate that the morning-time benefit for generalization was consistent across all bridge-satellite feature types.

**Generalization in Test 2.** In Session 2, there was no difference in generalization (functional judgments:  $t[28] = -0.52$ ,  $p = .61$ ; recognition:  $t[28] = 0.20$ ,  $p = 0.84$ ) between Morning and Evening groups, and neither group significantly differed from chance ( $p$ 's > .09; [S1 Fig](#)). There was a difference between groups in the change from Test 1 to Test 2 in functional judgment generalization ( $t[28] = -2.44$ ,  $p = .02$ ) and a marginal difference for recognition generalization ( $t[28] = -1.93$ ,  $p = .06$ ; [S2 Fig](#)). However, these differences cannot be interpreted as sleep effects because of the strong initial difference between groups (the time of day effect described above). Indeed, a multiple regression including initial performance and group as predictors revealed that initial performance negatively predicted the change in generalization performance across the delay ( $B = -0.48$ ,  $t = -2.52$ ,  $p = .02$ ) whereas group was not a significant predictor ( $B = 0.06$ ,  $t = 0.10$ ,  $p = .33$ ), nor was there an interaction ( $B = -0.17$ ,  $t = -0.89$ ,  $p = .38$ ). There was no difference between groups in the change from Test 1 to Test 2 in code name memory ( $t[28] = -1.14$ ,  $p = .26$ ).

**Discussion.** Although Morning and Evening groups differed in their change in generalization performance across the delay in a manner consistent with a benefit of sleep on generalization, this effect was most likely due to a strong time of day effect observed in the first session, whereby generalization was better in the morning than evening. Other measures of performance such as initial learning and exemplar memory did not differ between Morning and Evening groups, suggesting an effect specific to generalization (although the Morning group also tended to rate trained satellites as more functional).

## Experiment 2

Due to a time of day effect on initial generalization performance in Experiment 1, we were unable to cleanly assess the influence of sleep on generalization. To address this in Experiment 2, we attempted to match initial generalization performance by providing six blocks of bridge training to the Evening group and only one block to the Morning group. This experiment is thus unlike all other experiments, where participants trained to a criterion to ensure that they successfully learned the bridge satellites. Otherwise, the protocol was identical to Experiment 1.

**Training performance.** The manipulation was successful in changing bridge performance: There was a significant difference between groups on feature accuracy in the last block of bridge training ( $t[25] = -4.87$ ,  $p = .0005$ ; Morning:  $0.69 \pm 0.06$ ; Evening:  $0.98 \pm 0.01$ ). Providing the Evening group with six times as much training as the Morning group resulted in superior learning of the bridge satellite features. Other training phase results are presented in [S1 Table](#).

**Basic measures of performance in Test 1.** Mean slider ratings did not differ between Morning and Evening groups for trained satellites (functional:  $t[25] = -0.55$ ,  $p = .59$ , recognition:  $t[25] = -0.62$ ,  $p = .54$ ) or for satellites overall (functional:  $t[25] = 0.88$ ,  $p = 0.39$ ; recognition: ( $t[25] = 0.92$ ,  $p = .37$ ). There was also no difference between groups in code name memory ( $t[25] = -1.45$ ,  $p = .16$ ).

**Initial generalization performance in Test 1.** There was no difference between Morning and Evening groups on either measure of generalization (functional:  $t[25] = -0.79, p = .44$ ; recognition:  $t[25] = 0.65, p = 0.52$ ; Fig 2B). Neither group differed from zero on either generalization measure ( $p$ 's  $> .12$ ). Manipulating the amount of bridge exposure was thus successful in matching initial generalization performance between groups.

**Change in generalization across sessions.** With matched initial generalization performance, we next examined the change in performance across the delay. Counter to our prediction that 12 hours including sleep (Evening group) should facilitate generalization relative to 12 hours awake (Morning group), we found no difference between groups on the change in generalization performance between Test 1 and Test 2 on either measure of generalization (functional:  $t[25] = 1.61, p = .12$ ; recognition:  $t[25] = 0.30, p = .77$ ; S2 Fig). In Session 2 there was again no difference in generalization between groups (functional:  $t[25] = -0.50, p = .62$ ; recognition:  $t[25] = -0.81, p = .42$ ; S1 Fig). Neither group was above chance on either generalization measure ( $p$ 's  $> .33$ ). There was no effect of sleep on code name memory ( $t[25] = -1.44, p = .16$ ).

**Discussion.** We successfully matched initial generalization performance by providing the Evening group with much more bridge satellite training than the Morning group (who received very limited bridge-satellite training). This is unlike Experiment 1, where participants learned the bridge-satellites to a performance criterion. Counter to our prediction that a night of sleep should facilitate generalization relative to a day of wake, we found no evidence that sleep had any effect on generalization. Instead, the time of day effect observed in Experiment 1 is echoed in this experiment: even with six times as much bridge training, evening participants are unable to generalize above chance and no differently than the morning participants, who received much less training.

### Experiment 3

In this experiment, our primary aim was to replicate the time of day effect observed in Experiment 1. The only design change was that the delay period between Session 1 and Session 2 was 24 hours instead of 12. We tested after 24 hours to verify that a sleep effect would not emerge when time of day was matched across test sessions. From this experiment onwards, we also assessed chronotype using the Morningness-Eveningness Questionnaire (MEQ; [60]) to examine potential interactions between circadian preference and time of day effects. Basic descriptive statistics for each group for the MEQ are provided in S2 Table.

**Basic measures of performance in Test 1.** There was no difference between groups on mean slider ratings for trained satellites (functional:  $t[25] = 1.22, p = .23$ ; recognition:  $t[25] = 0.55, p = .58$ ). However, the Evening group endorsed satellites overall as less functional ( $t[25] = 4.20, p = .0003$ ; Morning:  $0.31 \pm 0.04$ ; Evening:  $0.06 \pm 0.04$ ) and gave lower recognition ratings ( $t[25] = 2.11, p = .04$ ; Morning:  $0.04 \pm 0.05$ ; Evening:  $-0.11 \pm 0.05$ ). Evening participants also exhibited marginally better code name memory for trained satellites than Morning participants ( $t[25] = -1.99, p = .06$ ; Morning:  $0.45 \pm 0.04$ ; Evening:  $0.57 \pm 0.05$ ).

**Time of day effects on generalization in Test 1.** The main goal of this experiment was to test whether the morning-time benefit for generalization would replicate. Indeed, we found a significant difference between Morning and Evening groups on functional judgment generalization ( $t[25] = 4.15, p = .0003$ ; Fig 2C). The Morning group exhibited reliably positive generalization for functional judgments ( $t[13] = 3.63, p = .003$ ) indicating that they rated bridge-consistent satellites as more functional than bridge-inconsistent ones. The morning-time benefit for generalization also extended to recognition generalization, with the Morning group showing increased false alarms to novel bridge-consistent satellites than the Evening group ( $t$

[25] = 2.65,  $p = .01$ ). The Morning group did not differ from zero on recognition generalization ( $t[13] = 1.02$ ,  $p = 0.33$ ). Notably, the Evening group demonstrated reliably negative generalization on both functional judgments ( $t[12] = -2.30$ ,  $p = .04$ ) and recognition ( $t[12] = -3.22$ ,  $p = .007$ ), indicating they were more likely to endorse novel bridge-inconsistent satellites as functional and having been seen before relative to novel bridge-consistent satellites. Time since awakening from sleep did not relate to either generalization measure in either group ( $p$ 's > .24).

Regressions with time of day and bridge satellite feature type (prototypical, some atypical, mostly atypical) revealed a main effect of time of day on both functional judgment generalization ( $B = 0.17$ ,  $t = 3.96$ ,  $p = .0002$ ) and recognition generalization ( $B = 0.13$ ,  $t = 2.92$ ,  $p = .005$ ). There was also a main effect of feature type with better functional judgment generalization for satellites that were made up of increasingly more atypical features ( $B = 0.11$ ,  $t = 2.14$ ,  $p = .04$ ). A similar marginal main effect of feature type was observed for recognition generalization ( $B = .11$ ,  $t = 1.93$ ,  $p = .06$ ). Interaction terms did not reach significance ( $p$ 's > .50).

To examine whether circadian preference modulated the observed time of day effects, we ran regression models where we let MEQ score and time of day interact as predictors for generalization. This yielded a significant main effect of time of day on the functional judgment measure of generalization ( $B = 0.19$ ,  $t = 5.15$ ,  $p = .00003$ ), but no interaction between MEQ and time of day ( $B = -0.004$ ,  $t = -1.29$ ,  $p = .21$ ). This suggests that there is an overall benefit for generalization in the morning that is not influenced by an individual's circadian preference. There was also a main effect of MEQ on generalization ( $B = 0.01$ ,  $t = 3.01$ ,  $p = .006$ ), with more strongly morning-type participants exhibiting better generalization. In the model for recognition generalization, there was a morning-time benefit for generalization ( $B = 0.14$ ,  $t = 2.73$ ,  $p = .01$ ), but no main effect of MEQ ( $B = 0.01$ ,  $t = 1.05$ ,  $p = .31$ ) nor an interaction ( $B = -0.01$ ,  $t = -0.54$ ,  $p = .59$ ). Although these findings are consistent with the possibility that circadian preference does not interact with the time of day effects on generalization, it is possible that we did not have sufficient sampling of extreme chronotypes for an influence of circadian preference to emerge in our sample (most of our sample were neutral chronotypes; [S2 Table](#)).

**Generalization in Test 2.** We next assessed differences across the 24h delay for the two groups. In Test 2, there was no difference between groups on either measure of generalization, and neither group was above chance ( $p$ 's > .62; [S1 Fig](#)). As groups did not differ from zero on the generalization index, the lack of a time of day effect in Test 2 may be due to overall forgetting across the 24h delay. Similar to the results in Experiment 1, there were reliable differences between groups in the change in generalization from Test 1 to Test 2 (functional:  $t[25] = -2.79$ ,  $p = .01$ , recognition:  $t[25] = -2.14$ ,  $p = .04$ ; [S2 Fig](#)). However, only initial performance ( $B = -0.70$ ,  $t = -3.11$ ,  $p = .005$ ), and not group ( $B = 0.03$ ,  $t = 0.43$ ,  $p = .67$ ), predicted the change in generalization across the delay (no interaction:  $B = -0.30$ ,  $t = -1.35$ ,  $p = .19$ ). There was no difference between groups in the change in code name memory between sessions ( $t[25] = 1.01$ ,  $p = .32$ ).

**Discussion.** Experiment 3 replicated the morning-time benefit for generalization observed in Experiment 1. Not only were morning participants better at generalizing the functional property to novel bridge-consistent satellites, but they were also more likely to have false memories of having seen these satellites before.

The Evening group exhibited reliably negative generalization in this experiment, meaning they tended to rate novel bridge-inconsistent satellites as more functional and more familiar than novel bridge-consistent satellites. This result may be consistent with non-monotonic accounts of memory plasticity [77, 78]: Moderate instead of strong activation of related satellites during the study of the bridge items, possibly due to higher inhibition in the evening than morning, could result in repulsion instead of integration of the bridged categories.

## Experiment 4

One benefit of sleep is its restorative effect, whereby participants who have slept recently exhibit enhanced memory encoding [79, 80]. Could the morning-time benefit for generalization be driven by the Morning group having slept more recently than the Evening group? In the above experiments, the lack of a relationship between time since awakening and performance in the Morning group provide some evidence contrary to this account. To more directly disentangle a restorative sleep account from a time of day account, we ran a nap study in the afternoon where half the participants had a 90-minute nap opportunity prior to the 4:00PM experiment (Nap group) and the other half did not nap (No Nap group). There was no delayed test in this experiment (or any of the following experiments).

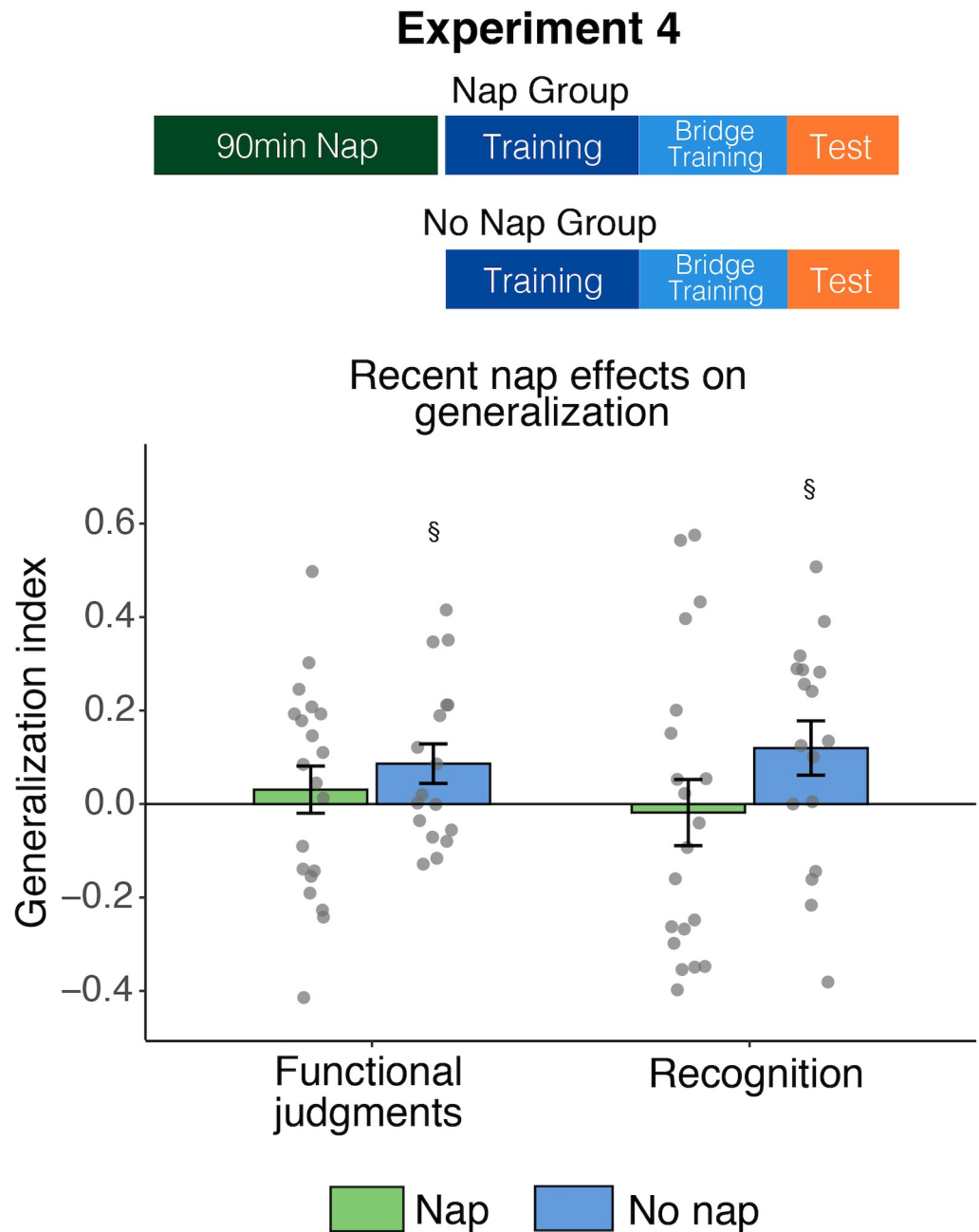
**Basic measures of performance and nap duration.** The mean total sleep time for the Nap group was 72.75 minutes (SE =  $\pm 4.09$ ). For the trained satellites, the Nap group provided marginally higher functional ratings ( $t[35] = 1.92, p = .06$ ; Nap:  $0.74 \pm 0.05$ ; No Nap:  $0.62 \pm 0.05$ ) and reliably higher recognition ratings ( $t[35] = 2.41, p = .02$ ; Nap:  $0.59 \pm 0.05$ ; No Nap:  $0.42 \pm 0.05$ ). Across all satellites, there was no difference between groups in mean slider judgments for recognition ( $t[35] = 1.22, p = .23$ ), although the Nap group rated them as marginally more functional ( $t[35] = 1.88, p = .07$ ; Nap:  $0.32 \pm 0.04$ ; No Nap:  $0.18 \pm 0.06$ ). There was no difference in code name memory between groups ( $t[35] = 1.00, p = .32$ ).

**Recent sleep effects on generalization.** Generalization did not differ between the Nap and No Nap groups (functional:  $t[35] = -0.83, p = .41$ ; recognition:  $t[35] = -1.47, p = .15$ ; Fig 3). The Nap group did not differ from zero on either measure of generalization (functional:  $t[19] = 0.61, p = 0.55$ , recognition:  $t[19] = -0.26, p = 0.80$ ). The No Nap group exhibited marginally positive generalization (functional:  $t[16] = 2.05, p = .06$ , recognition:  $t[16] = 2.06, p = .06$ ). A regression model with group and bridge satellite feature type (prototypical, some atypical, mostly atypical) predicting functional judgment generalization yielded no significant terms ( $p$ 's  $> .11$ ). For recognition memory generalization, the model revealed a marginal main effect of group ( $B = -0.07, t = -1.71, p = .09$ ), with slightly better generalization in the No Nap group, and a marginal main effect of bridge satellite feature type ( $B = 0.95, t = 1.92, p = .06$ ), with better generalization observed in satellites incorporating more atypical features (similar to Experiment 3). No other model terms were significant ( $p$ 's  $> .34$ ). Overall, the results suggest that having napped recently does not benefit generalization.

**Discussion.** Generalization performance did not differ between participants who had a 90-minute nap opportunity prior to the task relative to participants who did not nap. Therefore, the restorative effects of sleep following a nap [79, 80] do not appear to facilitate generalization in our paradigm. This finding is consistent with the lack of relationship in our earlier experiments between time since awakening in the morning and generalization. Thus, the morning-time benefit for generalization cannot be attributed to the Morning groups having slept recently, implicating a more general role of time of day. Counter to work demonstrating enhanced episodic learning following a nap [79, 80], initial learning measures and memory for exemplar names were also unaffected by having slept recently.

## Experiment 5

Experiments thus far demonstrated that generalization is better in the morning than evening (Experiments 1 and 3); that even with six times the amount of bridge satellite training, participants tested in the evening are no better at generalizing than participants tested in the morning (Experiment 2); and that this effect is not likely due to having slept recently (Experiment 4). Taken together, these experiments suggest that the time of day influences this form of generalization. How might generalization relate to other tasks known to be better at non-optimal



**Fig 3. Effects of having napped recently on generalization.** Generalization is depicted for a group that took an afternoon nap prior to starting the task and a group that did not nap. § $p < .1$ .

<https://doi.org/10.1371/journal.pone.0255423.g003>

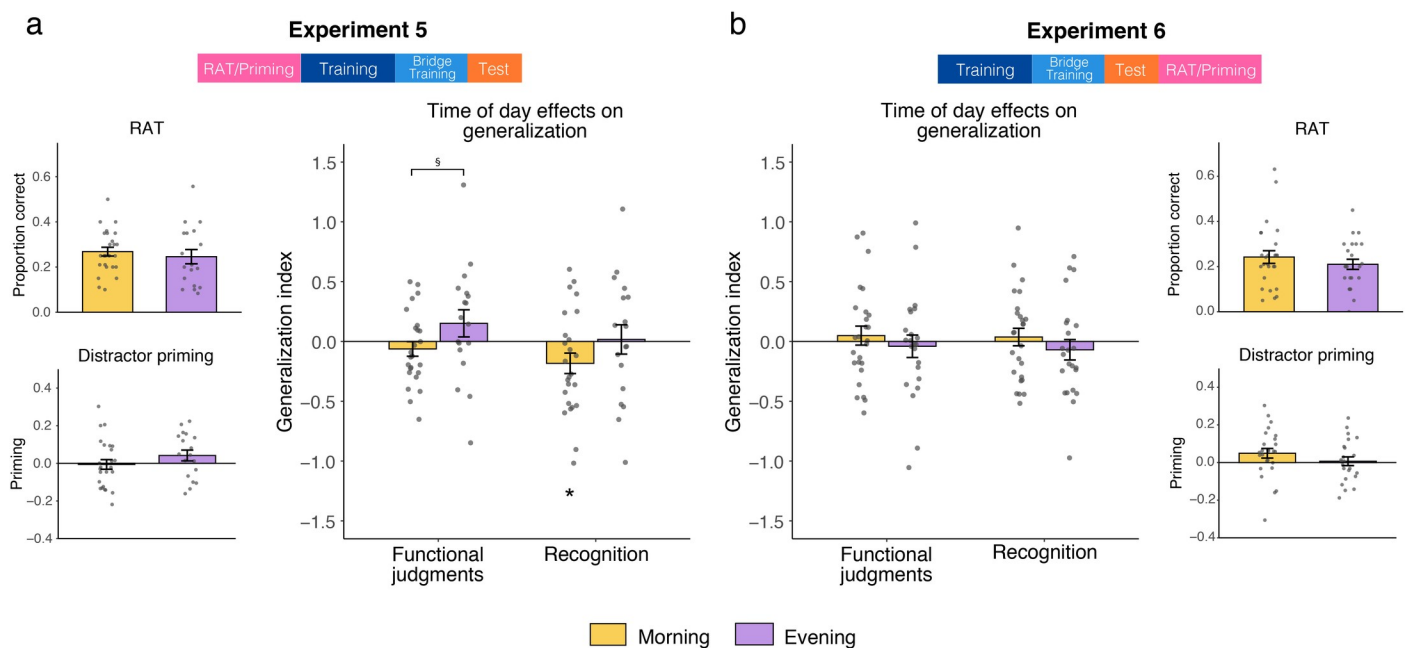
times of day, such as insight problem-solving on the Remote Associates Test (RAT; [63, 64] and memory for distractors [37, 50]? In this experiment, participants completed these tasks prior to the generalization paradigm. In order to keep the overall experiment length similar with the addition of these two tasks, we also used a simplified generalization paradigm with three instead of five exemplars per category, as well as fewer test trials. We hypothesized that there would be positive across-subject correlations between generalization and these additional tasks.

**Basic measures of performance.** The Evening group provided marginally higher functional ratings than the Morning group for trained satellites ( $t[41] = -1.93, p = .06$ ; Morning:  $0.47 \pm 0.05$ ; Evening:  $0.63 \pm 0.06$ ) and for recognition ( $t[41] = -1.86, p = .07$ ; Morning:  $0.45 \pm 0.04$ ; Evening:  $0.56 \pm 0.05$ ). The Evening group also rated satellites overall as marginally more functional than the Morning group ( $t[41] = -1.83, p = .07$ ; Morning:  $0.19 \pm 0.04$ ; Evening:  $0.32 \pm 0.06$ ; recognition:  $t[41] = -0.02, p = .99$ ). There was no difference between groups in code name memory ( $t[41] = -0.36, p = 0.72$ ). Overall, basic performance measures did not substantially differ between groups.

**Time of day effects on generalization.** Counter to our previous experiments, we found no effect of time of day on functional judgment generalization ( $t[41] = -1.79, p = .08$ ) and neither group differed from zero on this measure (Morning:  $t[24] = -1.03, p = .31$ , Evening:  $t[17] = 1.34, p = .20$ ; Fig 4A). Although not significant, the time of day effect was trending in the opposite direction as earlier experiments, with Evening participants exhibiting marginally better generalization than Morning participants. There was no difference in generalization between groups for recognition ( $t[41] = -1.38, p = .18$ ). While the Evening group was not significantly different from zero on generalization for recognition ( $t[17] = 0.14, p = .89$ ), the Morning group demonstrated negative generalization on this measure ( $t[24] = -2.14, p = .04$ ). Regression models with time of day and MEQ interacting as predictors of generalization yielded no significant terms ( $p$ 's  $> .11$ ), although there was a main effect of MEQ on recognition generalization ( $B = 0.02, t = 2.04, p = .05$ ), whereby greater morning preference was associated with better generalization.

**Time of day effects on RAT and distractor priming.** There was no difference between Morning and Evening groups on the RAT ( $t[41] = 0.63, p = .53$ ) or on the distractor priming test ( $t[41] = -1.22, p = .23$ ). A regression where we let time of day and MEQ score interact as predictors for each of these tasks yielded no significant terms ( $p$ 's  $> .22$ ).

To examine how the tasks relate, we ran correlations between performance on the RAT and distractor priming with measures of generalization separately for the Morning and Evening



**Fig 4. Time of day effects on RAT, distractor priming, and generalization with a simplified category structure.** Results for three tasks in (a) Experiment 5 and (b) Experiment 6, with different task orderings as indicated in the top schematics. \* $p < .05$ , \$ $p < .1$ .

<https://doi.org/10.1371/journal.pone.0255423.g004>



groups. None of these correlations were significant ( $p$ 's > .10). There was also no correlation between performance on the RAT and distractor priming task for either group ( $p$ 's > .15).

**Discussion.** In the current experiment, we found no evidence of the morning-time benefit for generalization observed in earlier experiments. The major change in this experiment was the use of a simplified generalization paradigm which had fewer exemplars per category. Given that exemplar variability is important for generalization [81–84], it is possible that this change weakened the strength of the learned category representations, thereby weakening generalization. Another possibility is that administering additional tasks interfered with how participants were performing the generalization task. This is possible given that previous work has shown certain tasks can push participants into more inhibitory states that linger into subsequent task performance [85]. We did not find any effect of time of day, or interaction with chronotype, on the RAT or on the distractor priming, perhaps due to our sample consisting mostly of neutral types (S2 Table). There were no correlations amongst any of these measures, suggesting that these tasks may be supported by different underlying mechanisms.

## Experiment 6

In Experiment 5, we did not replicate the time of day effect on generalization observed in previous experiments. This may have been due to the simplified category structure or to administering other tasks prior to the generalization task. To adjudicate between these possibilities, we ran the same paradigm used in Experiment 5, but administered the RAT and distractor priming test at the end of the study, after the generalization task.

**Basic measures of performance.** There was no difference between groups on the mean slider ratings provided for trained satellites (functional:  $t[47] = 0.26$ ,  $p = 0.80$ ; recognition:  $t[47] = 0.12$ ,  $p = 0.91$ ) or for satellites overall (functional:  $t[47] = 0.32$ ,  $p = 0.75$ ; recognition:  $t[47] = 0.48$ ,  $p = 0.64$ ). Code name memory also did not differ between groups ( $t[47] = 0.81$ ,  $p = .42$ ).

**Time of day effects on generalization.** As in Experiment 5, we found no significant difference between Morning and Evening groups on generalization (functional:  $t[47] = 0.73$ ,  $p = 0.47$ ; recognition:  $t[47] = 0.96$ ,  $p = 0.34$ ; Fig 4B). Neither group differed from zero on either generalization measure ( $p$ 's > .42). Regression models with time of day and MEQ interacting as predictors of generalization yielded no significant terms ( $p$ 's > .30).

**Time of day effects on additional measures.** There was no difference between Morning and Evening groups on the RAT ( $t[47] = 0.88$ ,  $p = .38$ ) or distractor priming ( $t[47] = 1.21$ ,  $p = 0.23$ ). Regression models where we let time of day and MEQ scores interact as predictors for performance on these tasks provided no significant terms ( $p$ 's > .27). No correlations between generalization and RAT or distractor priming were significant for the Morning or Evening group ( $p$ 's > .32). RAT and distractor priming were also not correlated with each other for either group ( $p$ 's > .14).

**Comparing the simplified and original paradigm.** The failure to replicate the clear time of day effect observed in earlier experiments suggests something may be different about learning in the simplified generalization paradigm. To investigate this, we examined how generalization and other performance measures differed as a function of paradigm type. For the following exploratory analyses, we combine Experiment 1 (E1) and E3 (original paradigm), and E5 and E6 (simple paradigm).

First, we assessed how the effect of time of day on generalization differs as a function of paradigm type. A two-way ANOVA revealed a significant interaction between paradigm type and time of day on functional judgment generalization ( $F[1,145] = 9.34$ ,  $p = .003$ ), but no main effect of paradigm type ( $F[1,145] = 0.50$ ,  $p = .48$ ) or time of day ( $F[1,145] = 2.45$ ,  $p = .12$ ).

Recognition generalization yielded similar results, with an interaction between paradigm type and time of day on generalization ( $F[1,145] = 4.25, p = .04$ ), and no main effects (paradigm type:  $F[1,145] = 0.40, p = .53$ , time of day:  $F[1,145] = 1.01, p = .32$ ). The significant interactions between paradigm type and time of day suggest that the morning-time benefit for generalization may only emerge in the context of strong category structure.

We next examined how other measures of performance varied as a function of paradigm type while ignoring time of day. As expected, given fewer exemplars to remember, the simple paradigm resulted in better code name memory than the original paradigm ( $t[147] = -2.90, p = .004$ ; original:  $0.50 \pm 0.02$ ; simple:  $0.59 \pm 0.02$ ). Although the mean recognition slider judgments for trained satellites did not differ ( $t[147] = 0.18, p = .85$ ), participants in the simple paradigm provided substantially higher recognition judgments for all satellites overall ( $t[147] = -4.42, p = .00002$ ; original:  $-0.05 \pm 0.03$ ; simple:  $0.09 \pm 0.02$ ). These findings suggest that the groups might have differed in their judgments for *novel* satellites. Indeed, we found that participants in the simplified paradigm provided higher recognition ratings for novel satellites ( $t[147] = -2.40, p = .02$ ; original:  $-0.25 \pm 0.04$ ; simple:  $-0.14 \pm 0.03$ ). This finding is likely to reflect the lack of atypical features in the novel test items for the simplified paradigm. Performance measures that did not vary as a function of paradigm include mean functional judgments for trained satellites and all satellites overall ( $p$ 's  $> .09$ ).

**Discussion.** As in Experiment 5, we did not observe a time of day effect on generalization in the simplified version of the paradigm. Thus, the failure to replicate earlier experiments appears related to the change in category structure, as opposed to the change in the order of tasks. Comparisons between the simplified and original paradigm revealed other basic differences in the way participants engaged with the task, with the simplified paradigm resulting in better memory for individual exemplars and a higher tendency to indicate recognition of novel satellites. We again found no time of day effect on the RAT or distractor priming task and no correlations between any of the tasks.

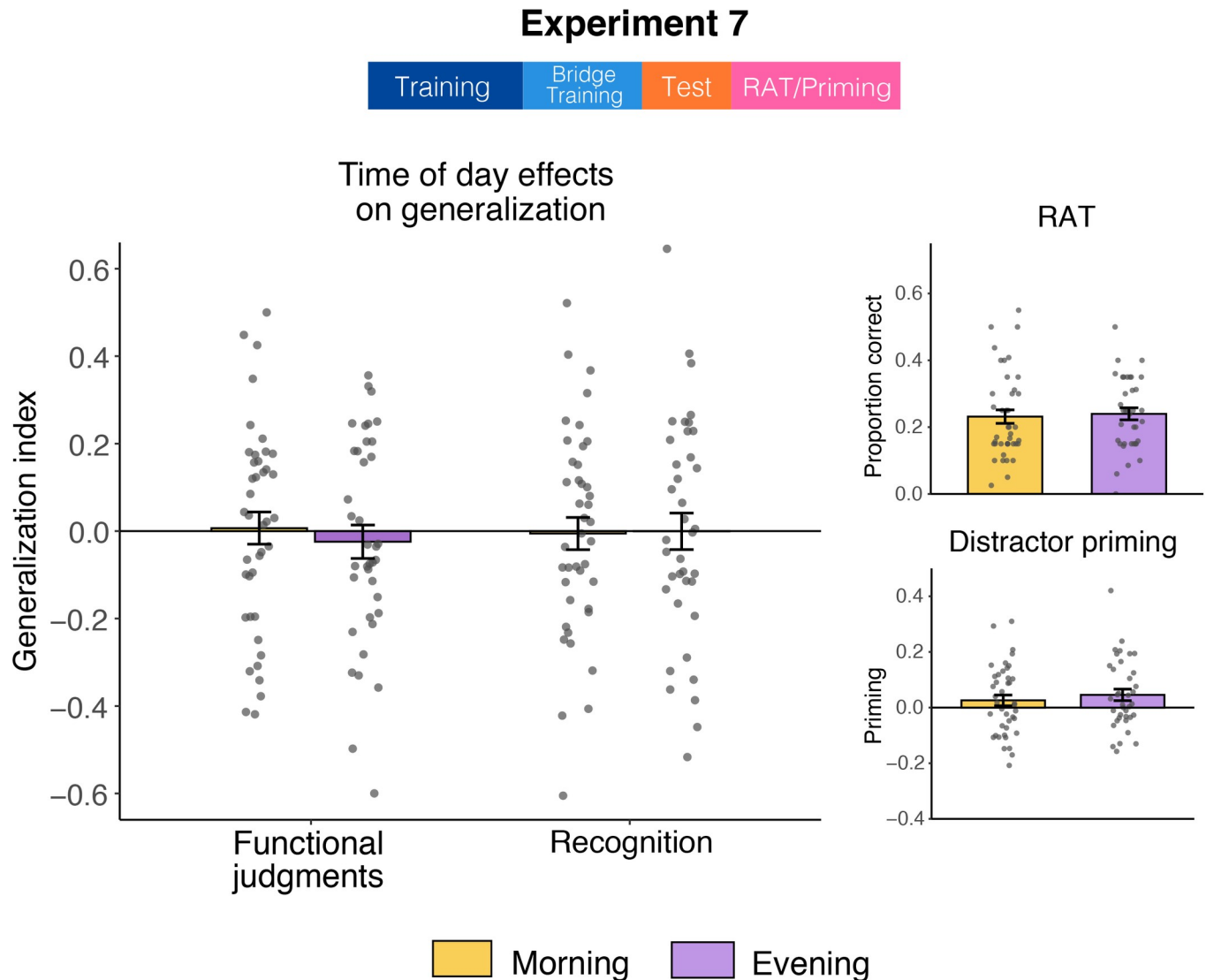
## Experiment 7

We did not observe a time of day effect in the context of simplified category structure, making it difficult to interpret the lack of relationships with the RAT and priming tasks. For the final experiment, we thus returned to the original generalization paradigm and administered the RAT and distractor priming task at the end of the experiment.

**Basic measures of performance.** Morning and Evening groups did not differ in ratings for trained satellites (functional:  $t[77] = 0.10, p = 0.92$ ; recognition:  $t[77] = 0.31, p = 0.76$ ) or for all satellites overall (functional:  $t[77] = 0.72, p = 0.47$ ; recognition: ( $t[77] = 0.34, p = .73$ ). There was no difference between groups in code name memory ( $t[77] = 1.46, p = .15$ ).

**Time of day effects on generalization.** There was no difference between Morning and Evening groups on either measure of generalization (functional:  $t[77] = 0.59, p = .56$ ; recognition:  $t[77] = -0.93, p = .93$ ; Fig 5). Neither group differed from zero on functional judgment generalization (Morning:  $t[40] = 0.19, p = .85$ ; Evening:  $t[37] = -0.63, p = .53$ ) or recognition generalization (Morning:  $t[40] = -0.15, p = .88$ ; Evening:  $t[37] = -0.01, p = .99$ ). MEQ did not interact with time of day in predicting either measure of generalization ( $p$ 's  $> .40$ ).

**Time of day effects on additional measures.** Morning and Evening groups did not differ on the RAT ( $t[76] = -0.30, p = .76$ ) or distractor priming task ( $t[76] = -0.70, p = 0.49$ ; Fig 5). Regression models where we let time of day and MEQ scores interact yielded no significant terms ( $p$ 's  $> .31$ ). Neither group showed a relationship between RAT and functional judgment generalization ( $p$ 's  $> .11$ ). The recognition memory measure of generalization exhibited a positive relationship with the RAT for the Morning group ( $r = 0.36, p = .02$ ) but not the Evening



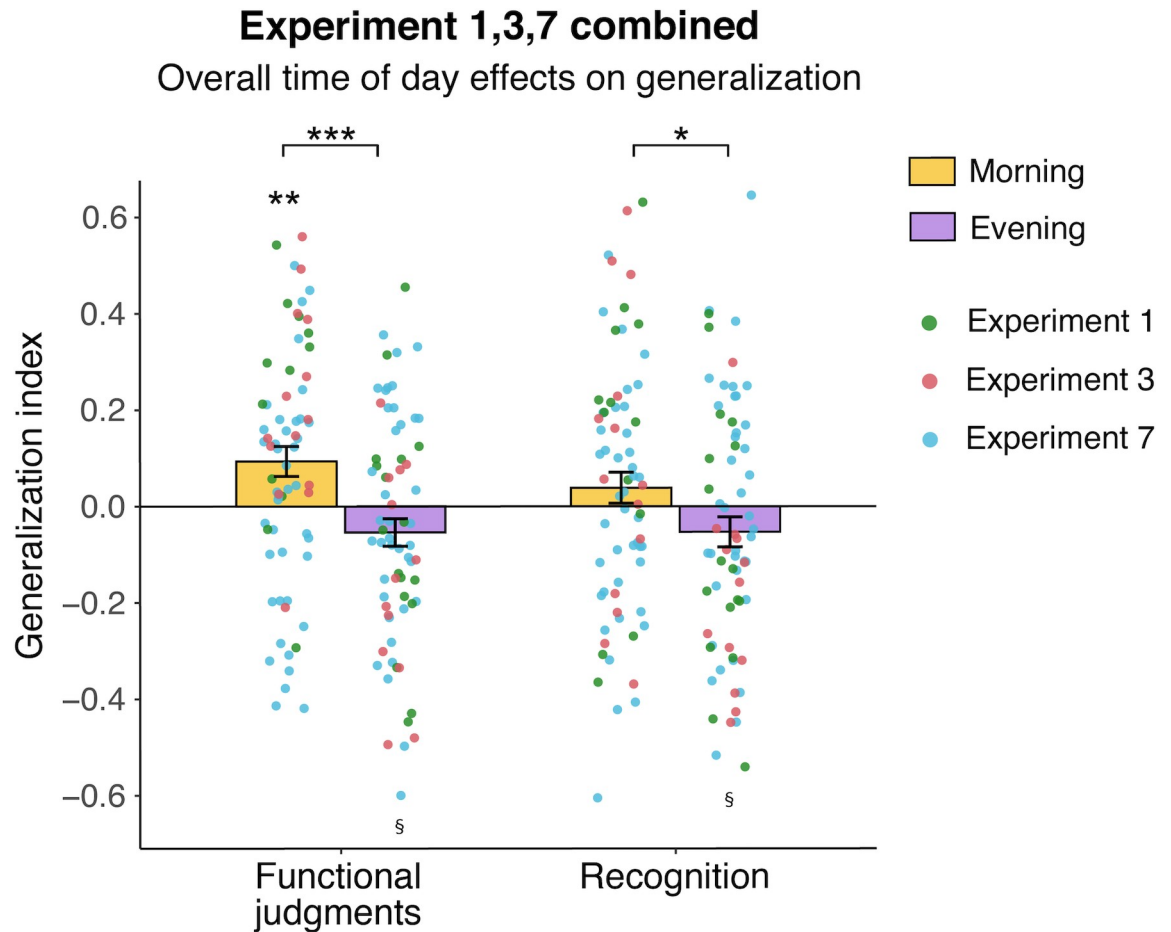
**Fig 5. Time of day effects in Experiment 7.** The generalization paradigm in this experiment was the same as Session 1 of Experiment 1 and Experiment 3. Additional measures, RAT and distractor priming, were administered after the generalization task, with performance shown on the right.

<https://doi.org/10.1371/journal.pone.0255423.g005>

group ( $r = -0.03$ ,  $p = .87$ ). This suggests that better performance on the RAT correlated with more generalization-related false memories, but only for the Morning group. Priming did not correlate with either measure of generalization for the Morning or Evening group ( $p$ 's  $> .27$ ). RAT and priming were also not correlated with each other for the Morning group ( $r = -0.08$ ,  $p = .61$ ), and there was a trending positive association for the Evening group ( $r = 0.28$ ,  $p = .09$ ).

**Discussion.** Despite returning to the original category structure, we did not find any evidence for a time of day effect on generalization, nor any evidence of generalization at all, in this experiment. In the following results section, we run a series of exploratory analyses to examine potential explanations.

We again did not find a time of day effect on the RAT or distractor priming task. In this experiment, we found that better insight problem-solving on the RAT was associated with more generalization-related false memories, but only for the Morning group. It is unclear how reliable this relationship is, given that it did not emerge in the previous experiments or in the



**Fig 6. Overall time of day effects on generalization in experiments with the original paradigm (Experiment 1 Test 1, Experiment 3 Test 1, Experiment 7).** Colored dots reflect individual participants from each experiment. \*\*\* $p < .001$ , \*\* $p < .01$ , \* $p < .05$ , \$ $p < .1$ .

<https://doi.org/10.1371/journal.pone.0255423.g006>

Evening group in this experiment. Nonetheless, it is worth noting that this finding is consistent with previous work showing that the RAT and false memory are positively associated [86].

**Across experiment analysis: Experiments 1, 3 and 7.** We employed the same generalization paradigm in Experiment 1 (E1), Experiment 3 (E3), and Experiment 7 (E7), yet we did not observe a time of day effect in E7. In this section, we pool data from these experiments to examine the time of day effect on generalization in a large sample size ( $n = 136$ ; 68 Morning, 68 Evening) and to examine performance differences across experiments that may account for the presence or absence of a time of day effect.

**Overall time of day effects on generalization.** We pooled the data from E7 and Session 1 from E1 and E3. Analyzed together ( $n = 136$ ), there is a clear effect of time of day on generalization, with better generalization in the morning ( $t[134] = 3.49$ ,  $p = .0007$ ; Fig 6). While the Morning group exhibited reliably positive generalization for functional judgments ( $t[67] = 3.01$ ,  $p = .004$ ), the Evening group demonstrated marginally negative generalization ( $t[67] = -1.88$ ,  $p = .06$ ). There was also a time of day effect on generalization for recognition, with more generalization-related false memories in the morning than evening ( $t[134] = 2.04$ ,  $p = .04$ ). Neither the Morning ( $t[67] = 1.20$ ,  $p = .23$ ) nor Evening group ( $t[67] = -1.70$ ,  $p = .09$ ) on their own were significantly different from zero for this measure, though the Evening group was trending towards negative generalization.

**Differences in time of day effects across experiments.** Despite identical generalization paradigms, E1 and E3 exhibited clear time of day effects whereas E7 did not. To assess how generalization in E7 differed from earlier experiments, we ran a two-way ANOVA on generalization with time of day (Morning, Evening) and experiment (E1, E3, E7) as factors. This revealed a main effect of time of day ( $F[1,130] = 13.34, p = .004$ ), with superior generalization in the morning compared to the evening. There was no main effect of experiment on generalization ( $F[2,130] = 2.01, p = .14$ ), but there was a significant interaction between time of day and experiment ( $F[2,130] = 6.37, p = .002$ ). Post-hoc t-tests revealed that the E7 Morning group's generalization was reliably worse than the Morning groups in E1 ( $t[52] = 3.16, p = .003$ ) and E3 ( $t[53] = 2.74, p = .008$ ). E1 and E3 Morning groups did not differ from each other ( $t[25] = 0.54, p = .59$ ). For the Evening groups, none of the experiments differed from any of the others ( $p's > .12$ ). Overall, this analysis suggests that it is specifically the Morning group that performed worse at generalization in E7 relative to the earlier experiments. The same two-way ANOVA above, but with the recognition measure of generalization, yielded similar effects (time of day:  $F[1,130] = 4.31, p = .04$ ; experiment:  $F[2,130] = 0.52, p = .60$ ; interaction:  $F[2,130] = 3.61, p = .03$ ). Post-hoc comparisons between groups did not yield any differences ( $p's > .10$ ), but the Evening group exhibited better generalization in E7 relative to E3 ( $t[49] = -2.32, p = .02$ ).

**Differences in other performance measures across experiments.** We next assessed differences in other measures of performance that might provide insight into why the time of day effect was not observed in E7. Generalization in E1 and E3 did not differ, so we collapsed them together (referred to now as E1&3) to compare them to E7. Designed as sleep studies, these earlier experiments had fewer participants than E7, so combining them also helps to better match sample size.

In the following two-way ANOVAs, we assess the effect of time of day and experiment, as well as the interaction between them, on various measures of performance. We identified several measures that differed across experiments that suggest differential task engagement. Participants in E7 ( $F[1,132] = 4.36, p = .04$ ), as well as evening participants ( $F[1,132] = 4.88, p = .03$ ) tended to rate satellites overall as less functional (no interaction:  $F[1,132] = 1.81, p = .18$ ). Furthermore, participants in E7 ( $F[1,132] = 4.27, p = .04$ ) rated the subset of *trained* satellites as less functional, which were the satellites that participants were explicitly told were functional. There was no main effect of time of day on the ability to rate trained satellites as functional ( $F[1,132] = 2.01, p = .16$ ). Although the interaction did not reach significance ( $F[1,132] = 2.56, p = .11$ ), we ran post-hoc t-tests to better understand the data. This revealed that the Morning group in E7 rated trained satellites as significantly less functional than the Morning group in E1&3 ( $t[66] = 2.44, p = .02$ ; E1&3:  $0.68 \pm 0.04$ ; E7:  $0.54 \pm 0.04$ ), but Evening groups across experiments did not differ from each other ( $t[66] = 0.37, p = .71$ ) (S3 Fig).

We also examined possible differences in slider strategy use by assessing variability in the overall slider responses (calculated as standard deviation) for each participant. For recognition judgments, participants in E7 exhibited reduced slider response variability ( $F[1,132] = 9.41, p = .003$ ). A significant interaction between experiment and time of day ( $F[1,132] = 6.77, p = .01$ ) revealed that the Morning group in E7 ( $t[66] = 4.01, p = .0002$ ), but not the Evening group ( $t[66] = 0.36, p = .72$ ), provided these less variable responses. For functional judgments, similar trends were observed but did not reach significance ( $p's > .13$ ). There were no main effects of time of day on either of these measures ( $p's > .13$ ). As participants were instructed to make use of the full range of the response slider, these analyses suggest participants in E7 were less engaged with the task.

Reduced engagement with the task in E7 might also manifest in response times. Indeed, median response times on the slider judgments were significantly faster in E7 than E1&3

(functional slider:  $F[1,132] = 14.43, p = .0002$ ; recognition slider:  $F[1,132] = 9.59, p = .002$ ). Time of day and interactions in these models were not significant ( $p$ 's  $> .23$ ). Basic t-tests comparing these response time measures between E7 and E1&3 further established that E7 participants were responding much faster on both slider judgments (functional slider:  $t[134] = 3.76, p = .0002$ ; recognition slider:  $t[134] = 3.04, p = .003$ ; S4 Fig).

**Accounting for trained functional judgments in time of day effects on generalization in E7.** On E7 data alone, we ran regressions where we let trained functional judgments and time of day interact in predicting generalization. None of the terms reached significance for the functional judgment generalization model ( $p$ 's  $> .19$ ), although there was a marginal interaction between time of day and functional judgments ( $B = -0.37, t = -1.71, p = .09$ ). For recognition generalization, there was a significant interaction between trained functional judgments and time of day in predicting generalization ( $B = -0.32, t = -2.93, p = .005$ ; S3 Fig). Notably, accounting for trained functional judgments, a time of day effect on generalization does emerge ( $B = 0.1, t = 2.71, p = .008$ ), with better generalization in the morning. Consistent with the interaction, there was no main effect of trained functional judgments on generalization ( $B = 0.07, t = 0.66, p = .51$ ). Post-hoc correlations revealed that trained functional judgments positively correlated with generalization for the Morning group ( $r = 0.44, p = .004$ ), but there was no significant relationship for the Evening group ( $r = -0.22, p = .18$ ). Though these analyses are exploratory, they suggest that the time of day effect on generalization (at least for the recognition memory measure) might emerge only when morning participants can make stronger functional judgments of the trained satellites. This may explain the lack of time of day effect in E7, as Morning participants in E7 provided weaker functional judgments for trained satellites relative to Morning participants in E1&3. While there were other measures that differed in E7 from E1&3 described above, none of these yielded significant terms in these models.

**Discussion.** Collapsing across experiments with matched paradigms ( $n = 136$ ), there is a clear morning-time benefit for generalization. A comparison across experiments revealed that participants in Experiment 7, relative to Experiments 1 and 3, appear less engaged with the task across multiple measures, including less variable slider responses, taking less time to consider responses, and a reduced ability to rate trained satellites as functional (which were explicitly defined to be functional). In Experiment 7, time of day effects on generalization (at least for recognition) only emerged in Morning participants who were able to effectively rate trained satellites as functional. Though exploratory, these analyses suggest that the lack of replication of the time of day effects in Experiment 7 may be due to this cohort of participants taking the task less seriously; they had faster and less variable responses, in addition to weaker functional judgments for trained satellites (which were defined for the participants to be functional, suggesting poor instruction-following).

## General discussion

Across seven experiments we examined the influence of time of day and sleep on generalization in a category learning paradigm that depends on the integration of previously separate memories. After learning about three distinct categories of novel “functional” satellites, participants briefly encountered bridge satellites that combined features from two of the initially learned categories. We then assessed how participants generalize judgments of functionality and recognition to never-before-seen bridge exemplars that similarly combine features from these two categories.

Counter to our expectations, we found no evidence that sleep facilitates generalization in this paradigm (Experiments 1–3). In lieu of a sleep effect, we found a time of day effect, with

better generalization in the morning than the evening (Experiments 1 and 3). Indeed, even when Evening participants were provided with extensively more bridge training than Morning participants, the Evening group performed no better than the Morning group (Experiment 2). This time of day effect did not appear to be due to the Morning group having slept more recently, as napping provided no benefit to subsequent generalization (Experiment 4). Taken together, these experiments provide evidence that time of day impacts this form of generalization. In subsequent experiments that implemented a simpler category structure, we did not observe the time of day effect (Experiments 5 and 6). We also observed no effect in a final experiment using the more complex paradigm, but found some evidence to suggest that this participant sample engaged less with the task (Experiment 7). The failure to replicate time of day effects in these experiments brings into question the replicability and generalizability of these time of day effects. Nonetheless, collapsing across the three studies with matched paradigms ( $n = 136$ ) there is a strong morning-time benefit for generalization, which we believe is worthy of further investigation.

### Evidence that generalization is better in the morning

Several cognitive processes are known to fluctuate with the circadian day-night cycle [23, 24]. Our study contributes to this literature by suggesting that generalization across concepts in a novel semantic domain may also be better in the morning than evening. Specifically, we found that participants tested in the morning were not only better able to generalize a property (whether a satellite was “functional”) to novel exemplars, but they also exhibited more generalization-related false memories, consistent with findings that generalization occurs alongside false memory formation [68, 69, cf. 70, 71]. Our findings thus illustrate the costs and benefits of generalization: the brain state in the morning that affords us the ability to make useful inferences can also result in false memory for new experiences consistent with one’s prior knowledge. We speculate that states of lower inhibition at non-optimal times of day (the morning for most young adults; [32, 34, 39, 59]) might highlight relatively weak connections between memories, facilitating both generalization and false memory. This is consistent with previous work demonstrating that at non-optimal times of day there is both an increased ability to find remote connections between words [63, 64] and increased memory interference [35].

In addition, we found some evidence that the Evening group exhibited a sort of *anti*-generalization: They were more likely to falsely recognize and to endorse as functional novel bridge-*inconsistent* exemplars. Under non-monotonic plasticity accounts [77, 78], this “repulsion” between concepts can occur when related memories are only moderately activated, as might occur in the evening in a state of higher inhibition. Notably, Retrieval-Induced Forgetting, a memory phenomenon that taps into these non-monotonic plasticity and repulsion processes [87, 88], is stronger at optimal times of day and this has been attributed to higher inhibition [48].

Although time of day effects on the retrieval of well-established semantic knowledge are mixed [34, 39, 54–59], two studies have found that individuals rely more on existing semantic knowledge to guide decision-making at non-optimal times of day [89, 90]. Our findings indicate that newly acquired semantic knowledge is also influenced by time of day, in a direction consistent with these studies (as the morning tends to be a non-optimal time for young adults [39, 45, 59]).

Outside of the circadian literature, unexpected morning-time benefits have been observed in sleep studies on tasks that may invoke similar generalization processes, such as category learning [19, 20], generalization of fear extinction memories [15, 21], and perhaps linguistic generalization [22]. There is thus a convergence of evidence of a morning-time benefit for generalization.

## Circadian preference did not modulate time of day effects on generalization

Although the morning is a non-optimal time for young adults at a group level, we did not find that the time of day effect was modulated by individual differences in circadian preference. A major caveat to this finding is that our sample consisted predominantly of neutral chronotypes (S2 Table), whereas studies reporting chronotype-related effects typically sample more densely from extreme morning and evening types [34, 46, 63, 64]. Though neutral type young adults often fail to show time of day effects [59], morning testing has been classified as a non-optimal time for both evening and neutral type young adults [35, 91], so it is likely the morning constituted a non-optimal time for the vast majority of our sample (we had very few morning types). To examine whether an evening benefit could emerge in this task, future work with strong morning types, which could be achieved by sampling older adults [39, 46], would be informative. Another limitation of our ability to speak to this question is that we did not collect chronotypes in Experiment 1 (as we did not originally foresee a time of day effect).

## Time of day effects may be sensitive to category structure and task engagement

We did not find a time of day effect on generalization in a simplified version of our task that involved fewer exemplars per category. Consistent with participants having superior exemplar memory in this version, it is possible that with fewer exemplars participants were simply memorizing the features of each satellite with less sensitivity to the shared structure of the categories, thereby weakening generalization. In support of this possibility, several category learning studies have demonstrated that exemplar variability is particularly important for generalization [cf. 70, 81–84]. Indeed, neither Morning nor Evening groups exhibited reliable generalization in the experiments with fewer exemplars. In sum, the finding suggests that the strength of time of day effects on generalization may depend on features of the task design and task complexity. Another possibility is that the simplified paradigm lacked adequate sensitivity to detect generalization during the test phase, which had less variable items and substantially fewer total trials.

We also did not observe a time of day effect in an experiment where participants appeared to be less engaged with the task, as evidenced by less variable slider responses and faster response times overall. Participants in this experiment also tended to rate trained satellites as less functional. As participants were explicitly told that trained satellites were by definition functional, this behavior suggests a lack of task engagement. Taking this variable into account in exploratory analyses partially recovers the morning-time benefit for generalization. It is possible that participants in this experiment were less invested or motivated given that they were paid less overall than participants in earlier experiments, which involved longer visits or visiting the lab on two occasions.

Several prior studies have demonstrated that factors of a task, such as whether the task is implicit or explicit, can modulate time of day effects [44, 46, 47, 49, 50, 53]. The mixed findings we report may reflect these kinds of influences, given that our paradigm is a complex category learning task that may lend itself to the use of multiple strategies.

## Sleep did not promote this form of generalization

Counter to work demonstrating that sleep facilitates generalization [9–11, 13–15, 92], we found no difference in the change in generalization ability across a night of sleep compared to a day awake when initial performance was matched (Experiment 2) or when taken into



account via regression (Experiment 1). Our findings thus contribute to the handful of studies demonstrating instances in which sleep does not facilitate the ability to generalize [3, 4, 12, 16].

Why does sleep fail to promote generalization in our paradigm? By providing minimal opportunity to learn the tested structure during training, we intended for our paradigm to require offline consolidation to generalize. However, the immediate above-chance generalization in our experiments suggests that offline processing is not always necessary for this form of generalization that requires the integration of previously separate memories. In support of this idea, an fMRI study found evidence of integrated neural representations in a period shortly after encountering “bridge” events that linked previously unrelated experiences [18]. Behavioral signatures of a similar form of integration have been linked to neural reinstatement of relevant memories during the bridge learning experiences themselves [17]. It thus appears that inferences dependent upon the integration of previously separate memories can stem from rapid changes in neural representations that occur largely online (during learning or immediately after). However, even in the groups that failed to show reliable evidence of generalization online, there was still no evidence for a benefit for sleep in our experiments.

### **Time of day effects did not emerge for the RAT or distractor priming task**

Prior work has demonstrated that other aspects of cognition, such as insight problem-solving on the RAT ([63, 64] and memory for distractions (distractor priming; [37, 50] are facilitated at chronotype-defined non-optimal times of day. In our experiments, we did not observe this interaction with chronotype, nor an effect of time of day, on either of these tasks. How might these findings be reconciled? Unlike the prior studies, our sample consisted mostly of neutral chronotypes, making it less sensitive to chronotype-related effects. Indeed, the RAT studies mentioned above did not observe any time of day effects in neutral chronotypes. In addition, a major procedural difference across studies concerns the time at which the evening tests were administered: we tested participants in the late evening (9pm), whereas much of the previous work test in the late afternoon (4–5:30pm [37, 64]). This testing time difference may thus contribute to the discrepant findings, especially given that circadian effects on cognition can be nonlinear across the day [24].

We also did not observe any reliable relationships between these tasks and generalization. A major limitation of this finding is that these measures were only collected in the set of generalization experiments that did not show time of day effects, making the lack of relationships difficult to interpret. Setting this issue aside, to the extent that these tasks are supported by different brain regions [28, 93, 94], a lack of a relationship does not preclude the possibility that they all rely on inhibition as a key mechanism, as circadian-related fluctuations in inhibition may be non-uniform across the brain [95, 96].

### **Possible neural mechanisms of generalization and their relationship to inhibition**

Although we have suggested that changes in inhibition may underlie the time of day effect on generalization, we did not assess neural inhibition directly in this study. An important avenue for future research will be to more directly establish how time of day changes in neural excitation and inhibition relate to behavioral outcomes. For instance, our research, combined with other behavioral evidence suggesting reduced inhibition in the morning, runs counter to the synaptic homeostasis hypothesis [97] whereby there is a gradual increase in cortical neural excitation across the day [98, 99] that is only renormalized by sleep (i.e., cortical inhibition is strongest in the morning). We also did not find that time since awakening in the morning, or an afternoon nap, related to generalization ability in our experiments. Instead, reduced

morning time inhibition may be better explained by separate circadian neural changes, such as the evening decrease in cortical excitability [100].

An important factor to consider is how circadian fluctuations in inhibition vary by brain region [95, 96]. Generalization in our category learning paradigm is likely dependent on the hippocampus [4, 67, 93, 101–105]. In rodents, hippocampal excitatory-inhibitory balance is strongly modulated by circadian rhythms [106–108] and neurotransmitter concentrations that alter this balance have been linked to circadian effects on hippocampal memory processes [109, 110]. Although we can currently only speculate as to how neural changes in inhibition relate to behavioral measures of inhibition and generalization in humans, neuroimaging tools that measure the concentrations of excitatory and inhibitory neurotransmitters in different brain regions [111–113] provide an exciting opportunity to investigate this more directly.

In addition, time of day is just one factor that influences inhibition. Changes in the excitatory-inhibitory balance in the brain vary dramatically across the lifespan [114, 115], across different clinical populations [116–118], across sleep stages [119–122], and as a function of task demands [85, 123]. Theories of how inhibition shapes the organization of memories [124–126] will thus be critical for a broad understanding, across many states and populations, of how memories interact to give rise to our ability to generalize [127]. This venture can also provide important insight into the putative neural mechanisms of generalization [67, 103, 104, 128, 129], as different theories make different predictions about the role of excitation and inhibition during memory formation and retrieval [67, 130].

### **Broader implications of time of day effects on measures and everyday cognition**

Our study serves as a cautionary tale for sleep designs that do not control for time of day—in particular, it is crucial to be able to verify that pre-delay behavior is matched between a sleep and wake group. Time of day effects may also be important to consider more broadly in experimental design and scheduling. For neuroimaging, significant influences of time of day have been observed in both functional [28, 131–134] and structural [135, 136] MRI scans.

Understanding how cognition fluctuates throughout the day is also critical for a better understanding and optimization of everyday behavior. Circadian rhythms interact with nearly every aspect of our daily lives, not least when we attend school or work [137, 138]. While various aspects of cognition are impaired in the morning for young adults [24, 39], leading this time to be considered “non-optimal,” the finding that generalization is improved in the morning in this population demonstrates that these effects may depend crucially on the kind of cognition in question. Characterizing these fluctuations and trade-offs can both allow us to better understand the mechanisms of cognition and to better match our lives to the ever-changing states of our brains.

### **Supporting information**

#### **S1 Table. Training performance.**

(PDF)

#### **S2 Table. Survey measures.**

(PDF)

#### **S1 Fig. Generalization in Test 2.**

(PDF)

**S2 Fig. Change in generalization across sessions (Test 2 generalization–Test 1 generalization).**

(PDF)

**S3 Fig. Functional judgments for trained satellites.**

(PDF)

**S4 Fig. Overall slider response times.**

(PDF)

**S1 File. Data files.**

(ZIP)

## Acknowledgments

We thank Lynn Hasher and Joan Ngo for sharing materials for the distractor priming task, and Kevin Madore for sharing materials for the Remote Associates Test.

## Author Contributions

**Conceptualization:** Robert Stickgold, Anna C. Schapiro.

**Formal analysis:** Marlie C. Tandoc, Mollie Bayda, Eileen Cho, Roy Cox, Anna C. Schapiro.

**Investigation:** Mollie Bayda, Craig Poskanzer, Eileen Cho, Anna C. Schapiro.

**Visualization:** Marlie C. Tandoc.

**Writing – original draft:** Marlie C. Tandoc.

**Writing – review & editing:** Marlie C. Tandoc, Mollie Bayda, Craig Poskanzer, Eileen Cho, Roy Cox, Robert Stickgold, Anna C. Schapiro.

## References

1. Landmann N, Kuhn M, Piosczyk H, Feige B, Baglioni C, Spiegelhalder K, et al. The reorganisation of memory during sleep. *Sleep Med Rev*. 2014; 18: 531–541. <https://doi.org/10.1016/j.smrv.2014.03.005> PMID: 24813468
2. Rasch B, Born J. About Sleep's Role in Memory. *Physiol Rev*. 2013; 93: 681–766. <https://doi.org/10.1152/physrev.00032.2012> PMID: 23589831
3. Mirković J, Gaskell MG. Does Sleep Improve Your Grammar? Preferential Consolidation of Arbitrary Components of New Linguistic Knowledge. *PLoS One*. 2016; 11: e0152489. <https://doi.org/10.1371/journal.pone.0152489> PMID: 27046022
4. Schapiro AC, McDevitt EA, Chen L, Norman KA, Mednick SC, Rogers TT. Sleep Benefits Memory for Semantic Category Structure While Preserving Exemplar-Specific Information. *Sci Rep*. 2017;7. <https://doi.org/10.1038/s41598-017-00035-9> PMID: 28127057
5. Durrant SJ, Taylor C, Cairney S, Lewis PA. Sleep-dependent consolidation of statistical learning. *Neuropsychologia*. 2011; 49: 1322–1331. <https://doi.org/10.1016/j.neuropsychologia.2011.02.015> PMID: 21335017
6. Dumay N, Gaskell MG. Sleep-Associated Changes in the Mental Representation of Spoken Words. *Psychol Sci*. 2007; 18: 35–39. <https://doi.org/10.1111/j.1467-9280.2007.01845.x> PMID: 17362375
7. Tamminen J, Payne JD, Stickgold R, Wamsley EJ, Gaskell MG. Sleep Spindle Activity is Associated with the Integration of New Memories and Existing Knowledge. *J Neurosci*. 2010; 30: 14356–14360. <https://doi.org/10.1523/JNEUROSCI.3028-10.2010> PMID: 20980591
8. McClelland JL, McNaughton BL, O'Reilly RC. Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychol Rev*. 1995; 102: 419–457. <https://doi.org/10.1037/0033-295X.102.3.419> PMID: 7624455

9. Ellenbogen JM, Hu PT, Payne JD, Titone D, Walker MP. Human relational memory requires time and sleep. *Proc Natl Acad Sci*. 2007; 104: 7723–7728. <https://doi.org/10.1073/pnas.0700094104> PMID: [17449637](https://pubmed.ncbi.nlm.nih.gov/17449637/)
10. Fenn KM, Nusbaum HC, Margoliash D. Consolidation during sleep of perceptual learning of spoken language. *Nature*. 2003; 425: 614–616. <https://doi.org/10.1038/nature01951> PMID: [14534586](https://pubmed.ncbi.nlm.nih.gov/14534586/)
11. Friedrich M, Wilhelm I, Born J, Friederici AD. Generalization of word meanings during infant sleep. *Nat Commun*. 2015; 6: 1–9. <https://doi.org/10.1038/ncomms7004> PMID: [25633407](https://pubmed.ncbi.nlm.nih.gov/25633407/)
12. Hennies N, Lewis PA, Durrant SJ, Cousins JN, Lambon Ralph MA. Time- but not sleep-dependent consolidation promotes the emergence of cross-modal conceptual representations. *Neuropsychologia*. 2014; 63: 116–123. <https://doi.org/10.1016/j.neuropsychologia.2014.08.021> PMID: [25174663](https://pubmed.ncbi.nlm.nih.gov/25174663/)
13. Lau H, Tucker MA, Fishbein W. Daytime napping: Effects on human direct associative and relational memory. *Neurobiol Learn Mem*. 2010; 93: 554–560. <https://doi.org/10.1016/j.nlm.2010.02.003> PMID: [20176120](https://pubmed.ncbi.nlm.nih.gov/20176120/)
14. Lau H, Alger SE, Fishbein W. Relational Memory: A Daytime Nap Facilitates the Abstraction of General Concepts. Chapouthier G, editor. *PLoS ONE*. 2011; 6: e27139. <https://doi.org/10.1371/journal.pone.0027139> PMID: [22110606](https://pubmed.ncbi.nlm.nih.gov/22110606/)
15. Pace-Schott EF, Germain A, Milad MR. Effects of sleep on memory for conditioned fear and fear extinction. *Psychol Bull*. 2015; 141: 835–857. <https://doi.org/10.1037/bul0000014> PMID: [25894546](https://pubmed.ncbi.nlm.nih.gov/25894546/)
16. Werchan DM, Gómez RL. Wakefulness (Not Sleep) Promotes Generalization of Word Learning in 2.5-Year-Old Children. *Child Dev*. 2014; 85: 429–436. <https://doi.org/10.1111/cdev.12149> PMID: [23962141](https://pubmed.ncbi.nlm.nih.gov/23962141/)
17. Chang CHC, Lazaridi C, Yeshurun Y, Norman KA, Hasson U. Relating the past with the present: Information integration and segregation during ongoing narrative processing. *J Cogn Neurosci*. 2021; 33: 1106–1128. [https://doi.org/10.1162/jocn\\_a\\_01707](https://doi.org/10.1162/jocn_a_01707)
18. Milivojevic B, Vicente-Grabovetsky A, Doeller CF. Insight Reconfigures Hippocampal-Prefrontal Memories. *Curr Biol*. 2015; 25: 821–830. <https://doi.org/10.1016/j.cub.2015.01.033> PMID: [25728693](https://pubmed.ncbi.nlm.nih.gov/25728693/)
19. Graveline YM, Wamsley EJ. The impact of sleep on novel concept learning. *Neurobiol Learn Mem*. 2017; 141: 19–26. <https://doi.org/10.1016/j.nlm.2017.03.008> PMID: [28288833](https://pubmed.ncbi.nlm.nih.gov/28288833/)
20. Lerner I, Lupkin SM, Corter JE, Peters SE, Cannella LA, Gluck MA. The influence of sleep on emotional and cognitive processing is primarily trait- (but not state-) dependent. *Neurobiol Learn Mem*. 2016; 134: 275–286. <https://doi.org/10.1016/j.nlm.2016.07.032> PMID: [27481220](https://pubmed.ncbi.nlm.nih.gov/27481220/)
21. Pace-Schott EF, Spencer RMC, Vijayakumar S, Ahmed NAK, Verga PW, Orr SP, et al. Extinction of conditioned fear is better learned and recalled in the morning than in the evening. *J Psychiatr Res*. 2013; 47: 1776–1784. <https://doi.org/10.1016/j.jpsychires.2013.07.027> PMID: [23992769](https://pubmed.ncbi.nlm.nih.gov/23992769/)
22. Xie X, Earle FS, Myers EB. Sleep facilitates generalisation of accent adaptation to a new talker. *Lang Cogn Neurosci*. 2018; 33: 196–210. <https://doi.org/10.1080/23273798.2017.1369551> PMID: [29372171](https://pubmed.ncbi.nlm.nih.gov/29372171/)
23. Krishnan HC, Lyons LC. Synchrony and desynchrony in circadian clocks: impacts on learning and memory. *Learn Mem*. 2015; 22: 426–437. <https://doi.org/10.1101/lm.038877.115> PMID: [26286653](https://pubmed.ncbi.nlm.nih.gov/26286653/)
24. Schmidt C, Collette F, Cajochen C, Peigneux P. A time to think: Circadian rhythms in human cognition. *Cogn Neurosci*. 2007; 24: 755–789. <https://doi.org/10.1080/02643290701754158> PMID: [18066734](https://pubmed.ncbi.nlm.nih.gov/18066734/)
25. Matchock RL, Mordkoff JT. Chronotype and time-of-day influences on the alerting, orienting, and executive components of attention. *Exp Brain Res*. 2009; 192: 189–198. <https://doi.org/10.1007/s00221-008-1567-6> PMID: [18810396](https://pubmed.ncbi.nlm.nih.gov/18810396/)
26. Valdez P, Ramírez C, García A, Talamantes J, Armijo P, Borrani J. Circadian rhythms in components of attention. *Biol Rhythm Res*. 2005; 36: 57–65. <https://doi.org/10.1080/09291010400028633>
27. Valdez P, Ramírez C, García A, Talamantes J, Cortez J. Circadian and homeostatic variation in sustained attention. *Chronobiol Int*. 2010; 27: 393–416. <https://doi.org/10.3109/07420521003765861> PMID: [20370477](https://pubmed.ncbi.nlm.nih.gov/20370477/)
28. Anderson JAE, Campbell KL, Amer T, Grady CL, Hasher L. Timing is everything: Age differences in the cognitive control network are modulated by time of day. *Psychol Aging*. 2014; 29: 648–657. <https://doi.org/10.1037/a0037243> PMID: [24999661](https://pubmed.ncbi.nlm.nih.gov/24999661/)
29. Barner C, Schmid SR, Diekelmann S. Time-of-day effects on prospective memory. *Behav Brain Res*. 2019; 376: 112179. <https://doi.org/10.1016/j.bbr.2019.112179> PMID: [31454674](https://pubmed.ncbi.nlm.nih.gov/31454674/)
30. García A, Ramírez C, Martínez B, Valdez P. Circadian rhythms in two components of executive functions: cognitive inhibition and flexibility. *Biol Rhythm Res*. 2012; 43: 49–63. <https://doi.org/10.1080/09291016.2011.638137>

31. Hahn C, Cowell JM, Wiprzycka UJ, Goldstein D, Ralph M, Hasher L, et al. Circadian rhythms in executive function during the transition to adolescence: the effect of synchrony between chronotype and time of day: Synchrony effects on executive function. *Dev Sci*. 2012; 15: 408–416. <https://doi.org/10.1111/j.1467-7687.2012.01137.x> PMID: 22490180
32. Hasher L, Zacks RT, May CP. Inhibitory control, circadian arousal, and age. *Attention and performance XVII: Cognitive regulation of performance: Interaction of theory and application*. Cambridge, MA, US: The MIT Press; 1999. pp. 653–675.
33. Lara T, Madrid JA, Correa Á. The Vigilance Decrement in Executive Function Is Attenuated When Individual Chronotypes Perform at Their Optimal Time of Day. van Wassenhove V, editor. *PLoS ONE*. 2014; 9: e88820. <https://doi.org/10.1371/journal.pone.0088820> PMID: 24586404
34. May CP. Synchrony effects in inhibitory control over thought and action. *J Exp Psychol Hum Percept Perform*. 1998; 363–379. <https://doi.org/10.1037//0096-1523.24.2.363> PMID: 9554091
35. Ngo KWJ, Hasher L. Optimal testing time for suppression of competitors during interference resolution. *Memory*. 2017; 25: 1396–1401. <https://doi.org/10.1080/09658211.2017.1309437> PMID: 28361617
36. Rothen N, Meier B. Time-of-day affects prospective memory differently in younger and older adults. *Aging Neuropsychol Cogn*. 2017; 24: 600–612. <https://doi.org/10.1080/13825585.2016.1238444> PMID: 27686115
37. Rowe G, Valderrama S, Centre B, Lenartowicz A. Attentional disregulation: a benefit for implicit memory. *Psychol Aging*. 2006; 826–830. <https://doi.org/10.1037/0882-7974.21.4.826> PMID: 17201503
38. West R, Murphy KJ, Armilio ML, Craik FIM, Stuss DT. Effects of Time of Day on Age Differences in Working Memory. *J Gerontol B Psychol Sci Soc Sci*. 2002; 57: P3–P10. <https://doi.org/10.1093/geronb/57.1.p3> PMID: 11773218
39. Yoon C, May CP, Hasher L. *Aging, Circadian Arousal Patterns, and Cognition. Cognition, Aging, and Self-Reports*. Abingdon, UK: Taylor & Francis; 1999. pp. 117–117. [https://doi.org/10.4324/9780203345115\\_chapter\\_6](https://doi.org/10.4324/9780203345115_chapter_6)
40. Coutanche MN, Koch GE, Paulus JP. Influences on memory for naturalistic visual episodes: sleep, familiarity, and traits differentially affect forms of recall. *Learn Mem*. 2020; 27: 284–291. <https://doi.org/10.1101/lm.051300.119> PMID: 32540918
41. Folkard S. Time of day and level of processing. *Mem Cognit*. 1979; 7: 247–252. <https://doi.org/10.3758/BF03197596>
42. Gerstner JR, Yin JCP. Circadian rhythms and memory formation. *Nat Rev Neurosci*. 2010; 11: 577–588. <https://doi.org/10.1038/nrn2881> PMID: 20648063
43. Hidalgo MPL, Zanette CB, Pedrotti M, Souza CM, Nunes PV, Chaves MLF. Performance of Chronotypes on Memory Tests during the Morning and the Evening Shifts. *Psychol Rep*. 2004; 95: 75–85. <https://doi.org/10.2466/pr0.95.1.75-85> PMID: 15460360
44. Himmer L, Müller E, Gais S, Schönauer M. Sleep-mediated memory consolidation depends on the level of integration at encoding. *Neurobiol Learn Mem*. 2017; 137: 101–106. <https://doi.org/10.1016/j.nlm.2016.11.019> PMID: 27913294
45. May CP, Hasher L, Stoltzfus ER. Optimal Time of Day and the Magnitude of Age Differences in Memory. *Psychol Sci*. 1993; 4: 326–330. <https://doi.org/10.1111/j.1467-9280.1993.tb00573.x>
46. May CP, Hasher L, Foong N. Implicit Memory, Age, and Time of Day: Paradoxical Priming Effects. *Psychol Sci*. 2005; 16: 96–100. <https://doi.org/10.1111/j.0956-7976.2005.00788.x> PMID: 15686574
47. Petros TV, Beckwith BE, Anderson M. Individual differences in the effects of time of day and passage difficulty on prose memory in adults. *Br J Psychol*. 1990; 81: 63–72. <https://doi.org/10.1111/j.2044-8295.1990.tb02346.x>
48. Pica G, Pierro A, Kruglanski AW. Effect of circadian rhythms on retrieval-induced forgetting. *Cogn Process*. 2014; 15: 29–38. <https://doi.org/10.1007/s10339-013-0575-z> PMID: 23836277
49. Puttaert D, Adam S, Peigneux P. Subjectively-defined optimal/non-optimal time of day modulates controlled but not automatic retrieval processes in verbal memory. *J Sleep Res*. 2019; 28: e12798. <https://doi.org/10.1111/jsr.12798> PMID: 30485575
50. Rothen N, Meier B. Time of day affects implicit memory for unattended stimuli. *Conscious Cogn*. 2016; 46: 1–6. <https://doi.org/10.1016/j.concog.2016.09.012> PMID: 27677048
51. Smarr BL, Jennings KJ, Driscoll JR, Kriegsfeld LJ. A time to remember: The role of circadian clocks in learning and memory. *Behav Neurosci*. 2014; 128: 283–303. <https://doi.org/10.1037/a0035963> PMID: 24708297
52. Winocur G, Hasher L. Age and time-of-day effects on learning and memory in a non-matching-to-sample test. *Neurobiol Aging*. 2004; 25: 1107–1115. <https://doi.org/10.1016/j.neurobiolaging.2003.10.005> PMID: 15212835

53. Yang L, Hasher L, Wilson DE. Synchrony effects in automatic and controlled retrieval. *Psychon Bull Rev.* 2007; 14: 51–56. <https://doi.org/10.3758/bf03194027> PMID: 17468781
54. Anderson MJ, Petros TV, Beckwith BE, Mitchell WW, Fritz S. Individual Differences in the Effect of Time of Day on Long-Term Memory Access. *Am J Psychol.* 1991; 104: 241. <https://doi.org/10.2307/1423157>
55. Dunne MP, Roche F, Hartley LR. Effects of Time of Day on Immediate Recall and Sustained Retrieval from Semantic Memory. *J Gen Psychol.* 1990; 117: 403–410. <https://doi.org/10.1080/00221309.1990.9921146> PMID: 28142342
56. Fabbri M, Mencarelli C, Adan A, Natale V. Time-of-day and circadian typology on memory retrieval. *Biol Rhythm Res.* 2013; 44: 125–142. <https://doi.org/10.1080/09291016.2012.656244>
57. Millar K, Styles BC, Wastell DG. Time of day and retrieval from long-term memory. *Br J Psychol.* 1980; 71: 407–414. <https://doi.org/10.1111/j.2044-8295.1980.tb01755.x>
58. Tilley A, Warren P. Retrieval From Semantic Memory at Different Times of Day. *J Exp Psychol Learn Mem Cogn.* 1983; 9: 718–724. <https://doi.org/10.1037/0278-7393.9.4.718> PMID: 6227684
59. May CP, Hasher L. Synchrony Affects Performance for Older but not Younger Neutral-Type Adults. *Timing Time Percept.* 2017; 5: 129–148. <https://doi.org/10.1163/22134468-00002087>
60. Horne JA, Östberg O. A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. *Int J Chronobiol.* 1976; 4: 97–110. PMID: 1027738
61. May CP. Synchrony effects in cognition: The costs and a benefit. *Psychon Bull Rev.* 1999; 6: 142–147. <https://doi.org/10.3758/bf03210822> PMID: 12199309
62. Hasher L, Chung C, May CP, Foong N. Age, time of testing, and proactive interference. *Can J Exp Psychol Can Psychol Expérimentale.* 2002; 56: 200–207. <https://doi.org/10.1037/h0087397> PMID: 12271750
63. Simor P, Polner B. Differential influence of asynchrony in early and late chronotypes on convergent thinking. *Chronobiol Int.* 2017; 34: 118–128. <https://doi.org/10.1080/07420528.2016.1246454> PMID: 27791394
64. Wieth MB, Zacks RT. Time of day effects on problem solving: When the non-optimal is optimal. *Think Reason.* 2011; 17: 387–401. <https://doi.org/10.1080/13546783.2011.625663>
65. Amer T, Campbell KL, Hasher L. Cognitive Control As a Double-Edged Sword. *Trends Cogn Sci.* 2016; 20: 905–915. <https://doi.org/10.1016/j.tics.2016.10.002> PMID: 27863886
66. Thompson-Schill SL, Ramscar M, Chrysikou EG. Cognition without control: When a little frontal lobe goes a long way. *Curr Dir Psychol Sci.* 2009; 18: 259–263. <https://doi.org/10.1111/j.1467-8721.2009.01648.x> PMID: 20401341
67. Kumaran D, McClelland JL. Generalization Through the Recurrent Interaction of Episodic Memories. *Psychol Rev.* 2012; 119: 573–616. <https://doi.org/10.1037/a0028681> PMID: 22775499
68. Carpenter AC, Schacter DL. Flexible Retrieval: When True Inferences Produce False Memories. *J Exp Psychol Learn Mem Cogn.* 2017; 43: 335–349. <https://doi.org/10.1037/xlm0000340> PMID: 27918169
69. Varga NL, Gaugler T, Talarico J. Are mnemonic failures and benefits two sides of the same coin?: Investigating the real-world consequences of individual differences in memory integration. *Mem Cognit.* 2019; 47: 496–510. <https://doi.org/10.3758/s13421-018-0887-4> PMID: 30617747
70. Bowman CR, Zeithamova D. Training set coherence and set size effects on concept generalization and recognition. *J Exp Psychol Learn Mem Cogn.* 2020; 46: 1442–1464 <https://doi.org/10.1037/xlm0000824> PMID: 32105147
71. de Araujo Sanchez MA, Zeithamova D. Generalization and source memory in acquired equivalence. *PsyArXiv [Preprint]* 2020. <https://doi.org/10.31234/osf.io/txjsh>
72. Johns MW. Reliability and factor analysis of the Epworth Sleepiness Scale. *Sleep.* 1992; 15: 376–381. <https://doi.org/10.1093/sleep/15.4.376> PMID: 1519015
73. Hoddes E, Zarcone V, Smythe H, Phillips R, Dement WC. Quantification of Sleepiness: A New Approach. *Psychophysiology.* 1973; 10: 431–436. <https://doi.org/10.1111/j.1469-8986.1973.tb00801.x> PMID: 4719486
74. Brainard DH. The Psychophysics Toolbox. *Spat Vis.* 1997; 10: 433–436. <https://doi.org/10.1163/156856897X00357> PMID: 9176952
75. Bowden EM, Jung-Beeman M. Normative data for 144 compound remote associate problems. *Behav Res Methods Instrum Comput.* 2003; 35: 634–639. <https://doi.org/10.3758/bf03195543> PMID: 14748508
76. Amer T, Hasher L. Conceptual Processing of Distractors by Older but Not Younger Adults. *Psychol Sci.* 2014; 25: 2252–2258. <https://doi.org/10.1177/0956797614555725> PMID: 25376192

77. Ritvo VJH, Turk-Browne NB, Norman KA. Nonmonotonic Plasticity: How Memory Retrieval Drives Learning. *Trends Cogn Sci*. 2019 [cited 13 Aug 2019]. <https://doi.org/10.1016/j.tics.2019.06.007> PMID: 31358438
78. Sievers B, Momennejad I. SAMPL: The Spreading Activation and Memory PLasticity Model. *Animal Behavior and Cognition*; 2019 Sep. <https://doi.org/10.1101/778563>
79. Mander BA, Santhanam S, Saletin JM, Walker MP. Wake deterioration and sleep restoration of human learning. *Curr Biol*. 2011; 21: R183–R184. <https://doi.org/10.1016/j.cub.2011.01.019> PMID: 21377092
80. Ong JL, Lau TY, Lee XK, van Rijn E, Chee MWL. A daytime nap restores hippocampal function and improves declarative learning. *Sleep*. 2020; zsa058. <https://doi.org/10.1093/sleep/zsaa058> PMID: 32227222
81. Baeck A, Maes K, Meel CV, Op de Beeck HP. The Transfer of Object Learning after Training with Multiple Exemplars. *Front Psychol*. 2016;7. <https://doi.org/10.3389/fpsyg.2016.00007> PMID: 26834680
82. Hahn U, Bailey TM, Elvin LBC. Effects of category diversity on learning, memory, and generalization. *Mem Cognit*. 2005; 33: 289–302. <https://doi.org/10.3758/bf03195318> PMID: 16028584
83. Perry LK, Samuelson LK, Malloy LM, Schiffer RN. Learn Locally, Think Globally: Exemplar Variability Supports Higher-Order Generalization and Word Learning. *Psychol Sci*. 2010; 21: 1894–1902. <https://doi.org/10.1177/0956797610389189> PMID: 21106892
84. Posner MI, Keele SW. On the genesis of abstract ideas. *J Exp Psychol*. 1968; 77: 353–363. <https://doi.org/10.1037/h0025953> PMID: 5665566
85. Radel R, Davranche K, Fournier M, Dietrich A. The role of (dis)inhibition in creativity: Decreased inhibition improves idea generation. *Cognition*. 2015; 134: 110–120. <https://doi.org/10.1016/j.cognition.2014.09.001> PMID: 25460384
86. Dewhurst SA, Thorley C, Hammond ER, Ormerod TC. Convergent, but not divergent, thinking predicts susceptibility to associative memory illusions. *Personal Individ Differ*. 2011; 51: 73–76. <https://doi.org/10.1016/j.paid.2011.03.018>
87. Hulbert JC, Norman KA. Neural Differentiation Tracks Improved Recall of Competing Memories Following Interleaved Study and Retrieval Practice. *Cereb Cortex*. 2015; 25: 3994–4008. <https://doi.org/10.1093/cercor/bhu284> PMID: 25477369
88. Norman KA, Newman EL, Detre G. A neural network model of retrieval-induced forgetting. *Psychol Rev*. 2007; 114: 887. <https://doi.org/10.1037/0033-295X.114.4.887> PMID: 17907868
89. Bodenhausen GV. Stereotypes as Judgmental Heuristics: Evidence of Circadian Variations in Discrimination. *Psychol Sci*. 1990; 1: 319–322. <https://doi.org/10.1111/j.1467-9280.1990.tb00226.x>
90. Yoon C. Age Differences in Consumers' Processing Strategies: An Investigation of Moderating Influences. *J Consum Res*. 1997; 24: 329–342. <https://doi.org/10.1086/209514>
91. Ngo KWJ, Biss RK, Hasher L. Time of day effects on the use of distraction to minimise forgetting. *Q J Exp Psychol*. 2018; 71: 2334–2341. <https://doi.org/10.1177/1747021817740808> PMID: 30362399
92. Werchan DM, Gómez RL. Generalizing memories over time: Sleep and reinforcement facilitate transitive inference. *Neurobiol Learn Mem*. 2013; 100: 70–76. <https://doi.org/10.1016/j.nlm.2012.12.006> PMID: 23257278
93. Schapiro AC, McDevitt EA, Rogers TT, Mednick SC, Norman KA. Human hippocampal replay during rest prioritizes weakly learned information and predicts memory performance. *Nat Commun*. 2018; 9: 3920. <https://doi.org/10.1038/s41467-018-06213-1> PMID: 30254219
94. Tik M, Sladky R, Luft CDB, Willinger D, Hoffmann A, Banissy MJ, et al. Ultra-high-field fMRI insights on insight: Neural correlates of the Aha!-moment. *Hum Brain Mapp*. 2018; 39: 3241–3252. <https://doi.org/10.1002/hbm.24073> PMID: 29665228
95. Bridi MCD, Zong F-J, Min X, Luo N, Tran T, Qiu J, et al. Daily Oscillation of the Excitation-Inhibition Balance in Visual Cortical Circuits. *Neuron*. 2020; 105: 621–629.e4. <https://doi.org/10.1016/j.neuron.2019.11.011> PMID: 31831331
96. Jasinska M, Pyza E. Circadian Plasticity of Mammalian Inhibitory Interneurons. *Neural Plast*. 2017; 2017: 1–12. <https://doi.org/10.1155/2017/6373412> PMID: 28367335
97. Tononi G, Cirelli C. Sleep and the Price of Plasticity: From Synaptic and Cellular Homeostasis to Memory Consolidation and Integration. *Neuron*. 2014; 81: 12–34. <https://doi.org/10.1016/j.neuron.2013.12.025> PMID: 24411729
98. Huber R, Mäki H, Rosanova M, Casarotto S, Canali P, Casali AG, et al. Human Cortical Excitability Increases with Time Awake. *Cereb Cortex*. 2013; 23: 1–7. <https://doi.org/10.1093/cercor/bhr363> PMID: 22275480

99. Lang N, Rothkegel H, Reiber H, Hasan A, Sueske E, Tergau F, et al. Circadian Modulation of GABA-Mediated Cortical Inhibition. *Cereb Cortex*. 2011; 21: 2299–2306. <https://doi.org/10.1093/cercor/bhr003> PMID: 21350047
100. Ly JQM, Gaggioni G, Chellappa SL, Papachilleos S, Brzozowski A, Borsu C, et al. Circadian regulation of human cortical excitability. *Nat Commun*. 2016; 7: 11828. <https://doi.org/10.1038/ncomms11828> PMID: 27339884
101. Bowman CR, Zeithamova D. Abstract memory representations in the ventromedial prefrontal cortex and hippocampus support concept generalization. *J Neurosci*. 2018; 2811–17. <https://doi.org/10.1523/JNEUROSCI.2811-17.2018> PMID: 29437891
102. Mack ML, Love BC, Preston AR. Building concepts one episode at a time: The hippocampus and concept formation. *Neurosci Lett*. 2018; 680: 31–38. <https://doi.org/10.1016/j.neulet.2017.07.061> PMID: 28801273
103. Whittington JC, Muller TH, Mark S, Chen G, Barry C, Burgess N, et al. The Tolman-Eichenbaum Machine: Unifying space and relational memory through generalisation in the hippocampal formation. *Cell*. 2019; 183: 1249–1263. <https://doi.org/10.1016/j.cell.2020.10.024> PMID: 33181068
104. Zeithamova D, Dominick AL, Preston AR. Hippocampal and Ventral Medial Prefrontal Activation during Retrieval-Mediated Learning Supports Novel Inference. *Neuron*. 2012; 75: 168–179. <https://doi.org/10.1016/j.neuron.2012.05.010> PMID: 22794270
105. Zeithamova D, Bowman CR. Generalization and the hippocampus: More than one story? *Neurobiol Learn Mem*. 2020; 175: 107317. <https://doi.org/10.1016/j.nlm.2020.107317> PMID: 33007461
106. Chaudhury D, Wang LM, Colwell CS. Circadian Regulation of Hippocampal Long-Term Potentiation. *J Biol Rhythms*. 2005; 20: 225–236. <https://doi.org/10.1177/0748730405276352> PMID: 15851529
107. Nakatsuka H, Natsume K. Circadian rhythm modulates long-term potentiation induced at CA1 in rat hippocampal slices. *Neurosci Res*. 2014; 80: 1–9. <https://doi.org/10.1016/j.neures.2013.12.007> PMID: 24406747
108. Raghavan AV, Horowitz JM, Fuller CA. Diurnal modulation of long-term potentiation in the hamster hippocampal slice. *Brain Res*. 1999; 833: 311–314. [https://doi.org/10.1016/s0006-8993\(99\)01523-1](https://doi.org/10.1016/s0006-8993(99)01523-1) PMID: 10375711
109. Ruby NF, Hwang CE, Wessells C, Fernandez F, Zhang P, Sapolsky R, et al. Hippocampal-dependent learning requires a functional circadian system. *Proc Natl Acad Sci*. 2008; 105: 15593–15598. <https://doi.org/10.1073/pnas.0808259105> PMID: 18832172
110. Snider KH, Sullivan KA, Obrietan K. Circadian Regulation of Hippocampal-Dependent Memory: Circuits, Synapses, and Molecular Mechanisms. *Neural Plast*. 2018; 7292540. <https://doi.org/10.1155/2018/7292540> PMID: 29593785
111. Puts NAJ, Edden RAE. In vivo magnetic resonance spectroscopy of GABA: A methodological review. *Prog Nucl Magn Reson Spectrosc*. 2012; 60: 29–41. <https://doi.org/10.1016/j.pnmrs.2011.06.001> PMID: 22293397
112. Frangou P, Correia M, Kourtzi Z. GABA, not BOLD, reveals dissociable learning-dependent plasticity mechanisms in the human brain. *eLife*. 2018; 7. <https://doi.org/10.7554/eLife.35854> PMID: 30355444
113. Koolschijn RS, Emir UE, Pantelides AC, Nili H, Behrens TEJ, Barron HC. The Hippocampus and Neocortical Inhibitory Engrams Protect against Memory Interference. *Neuron*. 2019; 101: 528–541.e6. <https://doi.org/10.1016/j.neuron.2018.11.042> PMID: 30581011
114. Gao F, Edden RAE, Li M, Puts NAJ, Wang G, Liu C, et al. Edited magnetic resonance spectroscopy detects an age-related decline in brain GABA levels. *NeuroImage*. 2013; 78: 75–82. <https://doi.org/10.1016/j.neuroimage.2013.04.012> PMID: 23587685
115. McQuail JA, Frazier CJ, Bizon JL. Molecular aspects of age-related cognitive decline: the role of GABA signaling. *Trends Mol Med*. 2015; 21: 450–460. <https://doi.org/10.1016/j.molmed.2015.05.002> PMID: 26070271
116. Heckers S, Konradi C. GABAergic mechanisms of hippocampal hyperactivity in schizophrenia. *Schizophr Res*. 2015; 167: 4–11. <https://doi.org/10.1016/j.schres.2014.09.041> PMID: 25449711
117. Nelson SB, Valakh V. Excitatory/Inhibitory Balance and Circuit Homeostasis in Autism Spectrum Disorders. *Neuron*. 2015; 87: 684–698. <https://doi.org/10.1016/j.neuron.2015.07.033> PMID: 26291155
118. Yizhar O, Fenno LE, Prigge M, Schneider F, Davidson TJ, O'Shea DJ, et al. Neocortical excitation/inhibition balance in information processing and social dysfunction. *Nature*. 2011; 477: 171–178. <https://doi.org/10.1038/nature10360> PMID: 21796121
119. Brown RE, Basheer R, McKenna JT, Strecker RE, McCarley RW. Control of sleep and wakefulness. *Physiol Rev*. 2012; 92: 1087–1187. <https://doi.org/10.1152/physrev.00032.2011> PMID: 22811426



120. Massimini M, Ferrarelli F, Huber R, Esser SK, Singh H, Tononi G. Breakdown of Cortical Effective Connectivity During Sleep. *Science*. 2005; 309: 2228–2232. <https://doi.org/10.1126/science.1117256> PMID: 16195466
121. Niethard N, Hasegawa M, Itokazu T, Oyanedel CN, Born J, Sato TR. Sleep-Stage-Specific Regulation of Cortical Excitation and Inhibition. *Curr Biol*. 2016; 26: 2739–2749. <https://doi.org/10.1016/j.cub.2016.08.035> PMID: 27693142
122. Tamaki M, Wang Z, Barnes-Diana T, Guo D, Berard AV, Walsh E, et al. Complementary contributions of non-REM and REM sleep to visual learning. *Nat Neurosci*. 2020; 1–7. <https://doi.org/10.1038/s41593-019-0563-4> PMID: 31844312
123. Shibata K, Sasaki Y, Bang JW, Walsh EG, Machizawa MG, Tamaki M, et al. Overlearning hyper-stabilizes a skill by rapidly making neurochemical processing inhibitory-dominant. *Nat Neurosci*. 2017; 20: 470–475. <https://doi.org/10.1038/nn.4490> PMID: 28135242
124. McKenzie S. Inhibition shapes the organization of hippocampal representations. *Hippocampus*. 2018; 28: 659–671. <https://doi.org/10.1002/hipo.22803> PMID: 28921762
125. Ramsaran AI, Schlichting ML, Frankland PW. The ontogeny of memory persistence and specificity. *Dev Cogn Neurosci*. 2019; 36: 100591. <https://doi.org/10.1016/j.dcn.2018.09.002> PMID: 30316637
126. Barron HC. Neural inhibition for continual learning and memory. *Curr Opin Neurobiol*. 2021; 67: 85–94. <https://doi.org/10.1016/j.conb.2020.09.007> PMID: 33129012
127. Robertson EM. Memory instability as a gateway to generalization. *PLOS Biol*. 2018; 16: e2004633. <https://doi.org/10.1371/journal.pbio.2004633> PMID: 29554094
128. Rogers TT, McClelland JL. *Semantic Cognition: A Parallel Distributed Processing Approach*. The MIT Press; 2004. <https://doi.org/10.7551/mitpress/6161.001.0001>
129. Shohamy D, Daw ND. Integrating memories to guide decisions. *Curr Opin Behav Sci*. 2015; 5: 85–90. <https://doi.org/10.1016/j.cobeha.2015.08.010>
130. Duncan KD, Schlichting ML. Hippocampal representations as a function of time, subregion, and brain state. *Neurobiol Learn Mem*. 2018; 153: 40–56. <https://doi.org/10.1016/j.nlm.2018.03.006> PMID: 29535044
131. Marek T, Fafrowicz M, Golonka K, Mojsa-Kaja J, Oginska H, Tucholska K, et al. Diurnal Patterns of Activity of the Orienting and Executive Attention Neuronal Networks in Subjects Performing a Stroop-Like Task: A Functional Magnetic Resonance Imaging Study. *Chronobiol Int*. 2010; 27: 945–958. <https://doi.org/10.3109/07420528.2010.489400> PMID: 20636208
132. Orban C, Kong R, Li J, Chee MWL, Yeo BTT. Time of day is associated with paradoxical reductions in global signal fluctuation and functional connectivity. Seymour B, editor. *PLOS Biol*. 2020; 18: e3000602. <https://doi.org/10.1371/journal.pbio.3000602> PMID: 32069275
133. Schmidt C, Collette F, Reichert CF, Maire M, Vandewalle G, Peigneux P, et al. Pushing the Limits: Chronotype and Time of Day Modulate Working Memory-Dependent Cerebral Activity. *Front Neurol*. 2015; 6. <https://doi.org/10.3389/fneur.2015.00199> PMID: 26441819
134. Vaisvilaite L, Hushagen V, Gronli J, Specht K. Time-of-day influence on resting-state fMRI: an overlooked factor contributing to the replication crisis? *BioRxiv [Preprint]*. 2020. <https://doi.org/10.1101/2020.08.20.258517>
135. Thomas C, Sadeghi N, Nayak A, Trefler A, Sarlls J, Baker CI, et al. Impact of time-of-day on diffusivity measures of brain tissue derived from diffusion tensor imaging. *NeuroImage*. 2018; 173: 25–34. <https://doi.org/10.1016/j.neuroimage.2018.02.026> PMID: 29458189
136. Trefler A, Sadeghi N, Thomas AG, Pierpaoli C, Baker CI, Thomas C. Impact of time-of-day on brain morphometric measures derived from T1-weighted magnetic resonance imaging. *NeuroImage*. 2016; 133: 41–52. <https://doi.org/10.1016/j.neuroimage.2016.02.034> PMID: 26921714
137. Alshabibi AS, Suleiman ME, Tapia KA, Brennan PC. Effects of time of day on radiological interpretation. *Clin Radiol*. 2020; 75: 148–155. <https://doi.org/10.1016/j.crad.2019.10.006> PMID: 31699432
138. Goldin AP, Sigman M, Braier G, Golombek DA, Leone MJ. Interplay of chronotype and school timing predicts school performance. *Nat Hum Behav*. 2020; 4: 387–396. <https://doi.org/10.1038/s41562-020-0820-2> PMID: 32042108