# Sleep consolidation of social memories and common ground

Yicheng Qiu*[1], Lewis V. Ball*[2], Catie Hopkins[1], Dan Denis[2], M. Gareth Gaskell[2], and Heather J. Ferguson[1]

[1]School of Psychology, University of Kent, Canterbury, UK

[2]School of Psychology, University of York, York, UK

* These authors contributed equally to the publication.

Correspondence to:

Heather Ferguson (h.ferguson@kent.ac.uk; https://orcid.org/0000-0002-1575-4820)

*Acknowledgements*

This work was carried out with the support of a grant to HF and GG from the Economic and Social Research Council (ESRC) (Ref: ES/X010643/1). We are grateful to Sarvani Dharwada, Barbara Bertozzi, Grace Day, Amy Howard, Jessica Neale, Joel Frimpong, Olivia Luckhurst, Sophie Shallow, Tessa Williams, Cristina Parera, and Devanshi Dagar for their contribution to data collection and conversation cleaning/transcribing.

*Abstract*

Sleep consolidation effects are well established in the cognitive domains of memory and language learning, however very little is known regarding sleep effects in social contexts. Across three experiments, we investigated whether sleep enhances the accuracy of socially formed memories and communicative efficiency, including the use of partner-specific common ground between interlocutors in conversation. Participants completed a computerised (Experiment 1) or face-to-face (Experiment 2) matcher-director task with two other people before and after 12 hours of sleep/wake. Both experiments showed more accurate item recognition and source memory after sleep vs. wake. The sleep group also used fewer words to describe old *vs.* new items in Experiment 2, suggesting an influence of sleep on communicative behaviour based on prior discourse, but this effect was not partner specific. In Experiment 3, the retention of item/source memory and common ground was measured across a two-hour, polysomnographically recorded nap. Results revealed no significant relationships between consolidation-linked sleep electrophysiology (e.g., slow oscillation-spindle coupling) and memory/common ground retention. Together, these findings indicate that sleep plays a key role in stabilising the social-cognitive foundations of communication but cannot elucidate the mechanism(s) underpinning this support. Regardless, by preserving conversational memories and the efficiency of shared understanding, sleep may help sustain smooth coordination in everyday interaction and social relationships.

*Keywords*: sleep; social interaction; memory; common ground; sleep spindles; slow oscillations

## Sleep consolidation of social memories and common ground

Face-to-face communication differs fundamentally from written communication because it requires not only understanding of words and phrases, but also sensitivity to the beliefs and knowledge shared with one's interlocutor. This process of adapting language to the listener is known as audience design (Clark & Murphy, 1982). According to the audience design hypothesis, speakers intentionally adapt their utterances to their conversational partners, for example by providing more detailed descriptions when a listener is unfamiliar with a referent. Through such coordination, interlocutors establish common ground (i.e. a representation of the knowledge, beliefs and experiences that they share; Clark & Brennan, 1991). Establishing common ground increases communication efficiency, for example by allowing speakers to use shorter or more specific descriptions once a referent has been mutually established (Clark & Wilkes-Gibbs, 1986; Hupet et al., 1993; Wilkes-Gibbs & Clark, 1992).

Common ground is not static; it can be partner-specific and adapts to the conversational context. If a new partner joins a dialogue, speakers often revert to more explicit, less efficient descriptions, reflecting absence of shared history with the new person (Horton & Gerrig, 2002; Isaacs & Clark, 1987; Schober & Clark, 1989; Yoon & Brown-Schmidt, 2014). Evidence suggests that interlocutors effectively track which objects or labels are shared with which partners (Gorman et al., 2013), relying on episodic memory to link conversational content to specific social partners.

The role of memory in common ground is further supported by generation effects, whereby speakers recall items in a conversation more accurately than listeners, consistent with the mnemonic advantage of self-produced information (Brown-Schmidt et al., 2023; Jacoby, 1978; Koriat et al., 1991; MacLeod et al., 2010; McKinley et al., 2017; Nault et al., 2023; Riefer et al., 2007; Slamecka & Graf, 1978). However, findings on source memory

(i.e., who said what, or with whom it was discussed) are less consistent, with some studies reporting poorer source memory for speakers than listeners (Brown et al., 1995; Jurica & Shimamura, 1999; Koriat et al., 1991, Exp 2) and others reporting no difference (Ahn & Brown-Schmidt, 2020; Knutsen et al., 2020; McKinley et al., 2017; Nault et al., 2023). This variability suggests that source memory for conversational partners is fragile and dependent on task and encoding conditions.

Episodic memory, traditionally distinguished from semantic memory by its role in encoding unique personal experiences (Dickerson & Eichenbaum, 2010), has been proposed as a key mechanism that supports partner-specific common ground (Brown-Schmidt et al., 2015; Horton, 2007; Horton & Gerrig, 2005). In this framework, conversational partners serve as retrieval cues for associated utterances, enabling the re-use of common ground. Studies of amnesia provide some support for these claims, since amnesic patients can successfully entrain labels with partners (Duff et al., 2008; Yoon et al., 2017), but show subtle deficits such as increased use of indefinite references (e.g., ‘a mug’ rather than ‘the mug’; Duff et al., 2011; Yoon et al., 2017), which indicates uncertainty about shared knowledge, and impaired management of partner-specific common ground in multi-party settings (Yoon et al., 2024). These patterns suggest that non-declarative memory systems can support some aspects of common ground (Brown-Schmidt & Duff, 2016; Duff et al., 2008), but that episodic memory, and hippocampal function in particular, are critical for its maintenance (Duff & Brown-Schmidt, 2012).

If common ground relies on episodic memory, then its formation and persistence should be shaped by factors known to modulate episodic memory. One such factor is sleep. Sleep is well established to protect newly formed episodic memories from interference (Berres & Erdfelder, 2021; Jenkins & Dallenbach, 1924; Yonelinas et al., 2019), and to actively consolidate them via hippocampal-cortical replay during non-rapid eye movement

(NREM) sleep (Born & Wilhelm, 2012; Diekelmann & Born, 2010; Klinzing et al., 2019; Rasch & Born, 2013). Coordinated interactions between slow oscillations, sleep spindles and hippocampal sharp-wave ripples are thought to facilitate this consolidation (Antony et al., 2018; Fernandez & Lüthi, 2020; Neske, 2016), and polysomnography studies consistently link spindle density and spindle-SO coupling with subsequent memory performance (Denis et al., 2021; Gais et al., 2002; Helfrich et al., 2018; Muehlroth et al., 2019; Ng et al., 2025; Schabus et al., 2004).

Despite this robust evidence for sleep's role in episodic memory consolidation, little is known about whether sleep also benefits socially-grounded memories, such as common ground. Yet there are reasons to expect such an effect (Diekelmann et al., 2018). Sleep has been shown to support memory for linguistic context, for example in word-meaning priming with homonyms (Gaskell et al., 2019; Mak et al., 2023), where sleep preserves context-specific meanings over time. More broadly, poor sleep has been linked to impairments in social cognition and relationships (Gordon & Chen, 2014; Maranges & McNulty, 2017; van der Helm et al., 2010; Zhang et al., 2019), which may be reflected in communication efficiency. Building on recent models of hippocampal involvement in language processing (Gaskell, 2024), one could argue that sleep enhances not only item memory but also the partner-specific contextual associations that are critical for common ground. In this paper, we directly test this hypothesis.

The current study uses a referential communication task to examine how sleep affects distinct memory processes underlying conversational behaviour. We do not consider common ground as a unitary episodic memory, but rather as a communicative outcome that is supported by multiple memory processes operating over different timescales. Specifically, item memory for the objects themselves indexes the retention of task content and may be supported by episodic encoding, however it is not inherently social. In contrast, source

memory (i.e. remembering who an item was discussed with) constitutes a partner-specific binding of content to social context, which more directly reflects episodic memory for a unique interaction. General reductions in description length for repeated items, and their maintenance over time, may reflect the abstraction of prior experiences into schema-like or item-specific representations rather than common ground. These are established within the conversation context of the task, but are not necessarily partner-specific. True partner-specific common ground involves adapting language differently for different interlocutors, which requires reinstating socially-specific details of the episode and relies on hippocampal-dependent episodic processes.

*The current study*

Across three experiments, we examined whether sleep enhances memory for conversational content and common ground, combining behavioural measures with polysomnography to assess the neural correlates of sleep-dependent consolidation. The general experimental design included two sessions and is illustrated in Figure 1. In Session 1, participants completed a referential communication task (McKinley et al., 2017), in which they took turns as "director" or "matcher" with two other participants to describe a set of images and establish common ground. Participants completed this task on adjacent computers in Experiment 1, and interacted face-to-face in Experiments 2 and 3. Next, participants completed a memory task assessing item and source memory from the referential communication task. This memory task was repeated at the start of Session 2 followed by a common ground production task. In Experiment 1, this was completed on a computer, and in Experiments 2 and 3, it was completed in a face-to-face referential communication task. The number of words used to describe items was used as a measure of common ground, and accuracy on the memory task was used as a measure of item and source memory. All

experiments included a battery of sleep measures to rule out any pre-existing differences between the wake and sleep groups.

In Experiments 1 and 2, sessions 1 and 2 were separated by a 12 hour delay. Participants in the "sleep" group completed Session 1 in the evening and returned the following morning after a night of sleep for Session 2. Participants in the "wake" group completed Session 1 in the morning and returned the same evening for Session 2 after a day awake. Experiment 1 tested whether item and source memory, as well as common ground, are better retained across a 12-hour delay that includes sleep compared to a delay with wakefulness, using a computer-mediated referential communication task. Experiment 2 extended Experiment 1 to face-to-face interactions, to examine whether the effects of sleep on memory and common ground generalise to a more naturalistic conversational setting. Experiment 3 tested whether a shorter, polysomnographically recorded nap is sufficient to produce sleep-dependent benefits for memory and common ground, and allowed us to link electrophysiological hallmarks of sleep (e.g., SO-spindle coupling) with memory performance. In this experiment, sessions 1 and 2 were separated by a 3-hour delay, and participants in the sleep group took a ~2 hour long, polysomnographically recorded nap in the lab in between sessions.

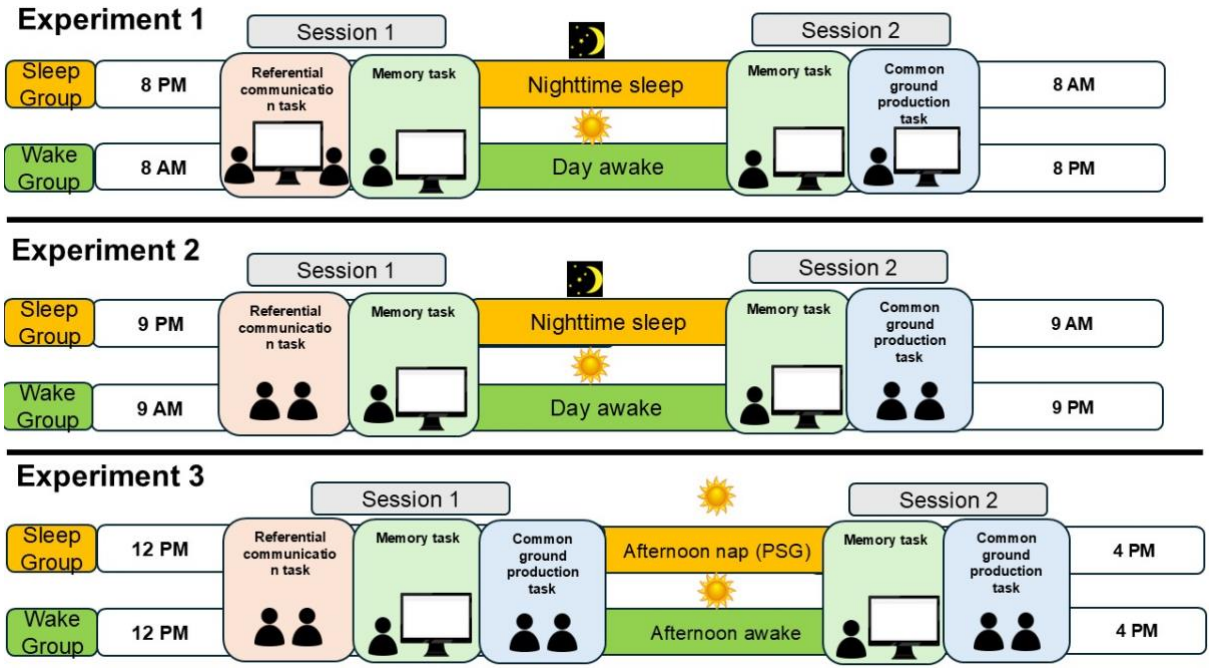


**Figure 1.** Schematic of the task procedure in Experiments 1-3.

*Experiment 1*

All hypotheses, methods and analysis plans were pre-registered on the Open Science Framework repository. The hypotheses for Experiment 1 and 2 were as follows (https://osf.io/ry4p7/overview?view_only=dd86d55d6d0e4d69b564a92971c1fbd0):

1. In Session 1, we expected to replicate previous work (e.g., McKinley et al., 2017) by finding that interlocutors establish a common ground that enables the use of entrained descriptions for the images (i.e., the number of words the 'director' uses to describe each image reduces from the first to the third trial in a given block).
2. In line with McKinley et al. (2017), we predicted that speakers (i.e., directors in the communication task) would have better item memory than listeners (i.e., matchers in the communication task), but their role in the original conversation would not impact source memory. In addition, we predicted that both item memory and source memory would be enhanced at Session 2 in the sleep group compared to the wake group.

3. We predicted that memory of conversation common ground would lead to shorter verbal descriptions for old *versus* new items, and shorter verbal descriptions for old items that are entrained and tested for the same person *versus* a different person. We predicted that these effects of common ground would be enhanced after 12 hours of sleep *versus* wake, as the sleep period supports better memory retention of the conversation and hence common ground.

## Methods

### *Participants*

The pre-registered sample size (N = 96) was based on that used in other sleep consolidation studies (e.g., Mak et al., 2024), and is comparable for each group to that used in McKinley et al. (2017, N = 55). To achieve this, a total of 108 participants were recruited from the University of [anonymised for peer-review] and completed the tasks in groups of four individuals. In case of participant no-shows at Session 1, a student research assistant filled in for the referential communication task, as in McKinley et al., 2017; their data did not contribute to any memory measures. Nine participants were excluded from the total sample because they failed to return for Session 2. The final sample therefore consisted of ninety-seven[1] participants (78 females and 19 males; $M_{age}$ = 19.52 years; *SD* = 1.69; 48 in wake group, and 49 in sleep group) who took part in the study across 35 groups.

Participants attended the experiment in the same group in two sessions, with a 12-hour lag between sessions to create two consolidation conditions: wake (Session 1 at 8am and Session 2 at 8pm, i.e., no sleep between two sessions) and sleep (Session 1 at 8pm and

[1] As a multi-person study requiring four participants in each session, we slightly exceeded the target sample size in all three experiments.

Session 2 at 8am, i.e., overnight sleep between two sessions). All participants were aged over 18, a native English speaker, and had normal or corrected-to-normal eyesight. Participants self-reported no neurological or developmental disorders (e.g., autism), no diagnosed sleep disorders, and were not taking any psychiatric or sleep-related medication during the time of the experiment. Participants were instructed not to consume any alcohol or take a nap (for Wake group) between the sessions; compliance was confirmed at the start of Session 2. Participants received course credits or a cash payment after completing the experiment. All participants read an information sheet and gave consent to participate. Ethics approval was obtained from the School of Psychology's Ethics Committee at the University of [anonymised for peer-review] (202317030908018859).

*Sleep measures*

Sleep quality was measured using a battery of established self-report questionnaires. This included the Pittsburgh Sleep Quality Index (PSQI, Buysse et al., 1989), the Stanford Sleepiness Scale (SSS, Hoddes et al., 1972), the Epworth Sleepiness Scale (ESS, Johns, 1991), and the reduced version of the Morningness-Eveningness Questionnaire (rMEQ, Adan & Almirall, 1991). The PSQI provides both quantitative and qualitative information about sleep over the past month and allows the calculation of component scores as well as a single global score. The SSS is a 7-point scale that assess subjective alertness throughout the day, while the ESS asks about the likelihood of dozing in specific situations. In other words, the SSS emphasizes general feelings of sleepiness, while the ESS focuses on situational sleepiness. The rMEQ is a shortened version of the MEQ (Horne & Ostberg, 1976) and assesses an individual's chronotype. Because the PSQI, ESS and rMEQ measure trait-like, long-term sleep behaviour, they were administered in Session 1 only. The SSS measures

state-like sleepiness, which can fluctuate over time, and was therefore administered in both sessions.

*Referential communication task*

The referential communication task aimed to establish common ground between interlocutors, and replicated the general procedure adopted in McKinley et al. (2017). Participants completed this task in pairs on adjacent computers in a shared lab and were matched with two different partners across four blocks. They sat approximately two metres apart, so that they would not see each other's screens but could still communicate with each other. Within each block/pairing, one participant was assigned the role of "director" and the other as "matcher" (see Figure 2).

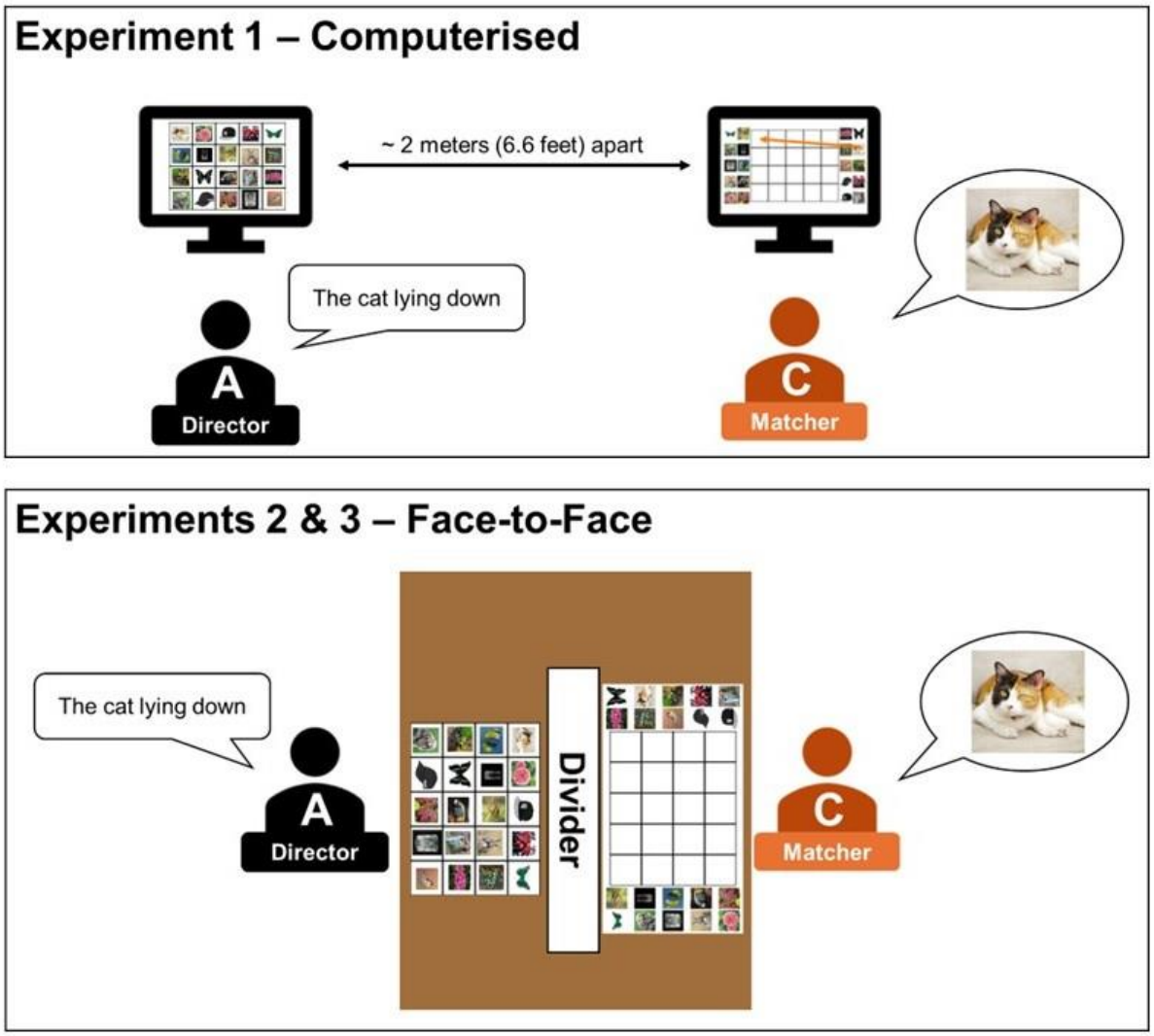


**Figure 2.** Layouts of Referential Communication Task (in Experiments 1-3).

The director viewed a 5 × 4 grid consisting of 20 different images depicting everyday objects and animals, and the matcher viewed an empty 5 × 4 grid with the same items located as clickable objects around the perimeter of the grid. The director's task was to provide verbal descriptions of each item to allow the matcher to identify the same item and place it in the corresponding (empty) square on their grid (using the mouse to click and drag). The conversations between directors and matchers were recorded using a microphone. Each block was made up of three trials involving the same 20 items but with a different spatial arrangement of items across trials (80 different items were therefore presented across the

whole task). The 20 items within a block were made up of ten different categories (leaves, fish, butterflies, drinking containers, flowers, frogs, cats, hats, rabbits and birds), with each category represented by two different exemplars. This meant that the director's description of each item had to provide enough detail to distinguish it from the other category member (see McKinley et al., 2017). To counterbalance the images across conditions (old *vs.* new), two lists of 160 images were used, with a different set of 80 images shown in the referential communication task across lists. These two lists were further counterbalanced to test a different set of 40 old/new images in Session 1 and 2 memory tasks.

Across the four blocks, participants completed each role twice, once with each partner. This enabled the development of partner-specific common ground. Common ground was measured in this task as the total number of words used by the director to enable the matcher to identify each item for each role/pair combination. Audio recordings of the conversations during the referential communication task were automatically transcribed using Qualtrics built-in transcription tool and then manually checked (and corrected for errors if necessary) by research assistants. As in McKinley et al. (2017), the transcripts were used to count the total number of words used to identify each item, including all descriptive words and phrases as well as any lexical disfluencies (e.g., "um", "uh") and location information (e.g., "on the top right is"). Total word count did not include feedback from the matcher (e.g., "yes", "OK") or task-irrelevant words (e.g., "Are you ready?").

*Memory task*

Participants completed this task individually on a computer. On each trial, participants viewed an image of an item and judged whether it was old or new (i.e., whether they recognised the item from the referential communication task or not). They then judged who they saw the item with (i.e., partner B or C; real first names were used), or to guess who they

might have seen the item with if they reported the item as new. This is to prevent introducing any incentive for participants to respond "new" to every item and has been an established procedure in memory research. This task was completed in both Session 1 and 2. The task included 80 images in total (40 old/40 new), with a different set of 80 old/new images shown in each session.

*Common ground production task*

Participants completed this task individually on a computer. On each trial, participants viewed an image of an item and were instructed to type into a box how they would describe this item to partner B or C (real partner names were used). This task was completed in session 2 only. The task included 80 images in total (40 old/40 new). Of the old images, 20 were discussed with partner B in the referential communication task in Session 1, and 20 were discussed with partner C, and half of each were probed by a cue to the same partner as the earlier task (source matching) and half were probed with the other partner (source mismatching). Common ground in this task was indexed by the total number of words used to describe each item in each condition.

*Procedure*

In Session 1, participants completed a questionnaire hosted on Qualtrics, which included demographic information (gender, age, and ethnicity) and the sleep quality battery (PSQI, ESS, SSS, and rMEQ). Then, participants were randomly paired up and rotated around two adjacent labs to complete the referential communication task, as described above. Following the referential communication task, they returned to individual computers to complete the memory task, which was programmed in PsychoPy. Twelve hours later, the same participants returned to the lab to complete Session 2, either after overnight sleep or daytime wake. In

Session 2, participants first completed the SSS questionnaire, then the memory task, and the common ground production task. Across all of our three experiments, participants were not explicitly informed that their memory would be tested at any stage.

## Results

### *Sleep measures and manipulation check*

The average sleep time reported by participants (over the last month, in the PSQI) was 1:10 AM, and the average wake time was 9:40 AM. A series of linear mixed-effect models were used to examine whether there were any differences in sleep quality measures between two groups. Participant was included as a random intercept. No significant group differences were found on any sleep measure (PSQI: $\beta$ = -0.76, *SE* = 0.71, *t* = -1.15, *p* = .252; ESS: $\beta$ = -0.11, *SE* = 0.71, *t* = -0.16, *p* = .874; rMEQ: $\beta$ = -0.43, *SE* = 0.65, *t* = -0.67, *p* = .506; SSS, Session 1: $\beta$ = 0.35, *SE* = 0.27, *t* = 1.32, *p* = .189; SSS, Session 2: $\beta$ = 0.12, *SE* = 0.27, *t* = 0.44, *p* = .659). This showed that the sleep patterns did not differ between groups.

### *Referential communication task*

The extent to which pairs of participants established common ground was indexed by the change in description length of the images from the first to the third trial (McKinley et al., 2017; Nault et al., 2023). To assess this in our experiment, we constructed a linear mixed-effect model in R using the *lme4* package (Bates et al., 2015). The dependent variable was the number of words used to describe each image, centred. Trial order was centred and entered as a fixed effect (Trial 1: -1, Trial 2: 0, and Trial 3: 1). Random effects included participant, item, and list (1/2). The random effects structure was constructed following Barr et al.'s (2013) recommendation, starting with a full model that contained by-participant, by-item and by-list random intercepts, as well as random slopes for the effect of Trial order for all three

random intercepts. We then used the 'buildmer' package in R (2.11, Voeten, 2023) to identify the maximal model, which was revealed to include random intercepts of subject and item. A BOBYQA optimizer with a set maximum of 200,000 iterations was used to increase chances of convergence. This model building procedure was applied to all mixed-effect models in this paper.

This analysis revealed a negative effect of trial centred, suggesting a linear decrease of utterance length from Trial 1 to Trial 3 ($\beta$ = -2.58, $SE$ = 0.04, $t$ = -66.93, $p < .001$). In other words, utterance length decreased successively across trials, with directors using 5.16 fewer words to describe the same image from Trial 1 to 3 (see Table 1).

**Table 1.** Summary of the average number of words used to describe each item in trials 1, 2 and 3 of the referential communication tasks across Experiments 1-3. Standard deviations are in parentheses. Note that Experiment 3 data is from the sleep group only ($n$ = 39).

| | **Experiment 1** | **Experiment 2** | **Experiment 3** |
|---|---|---|---|
| Trial 1 | 12.14 (7.22) | 13.01 (5.99) | 13.10 (6.61) |
| Trial 2 | 8.20 (4.94) | 8.40 (3.96) | 9.24 (4.46) |
| Trial 3 | 6.98 (4.15) | 7.02 (3.36) | 7.90 (3.88) |

*Memory task*

Our pre-registered analysis of the memory task analysed trial-level data using a logit mixed-effect model. This allowed us to test for effects of group and role on memory whilst remaining consistent with the analytic approaches in related memory and common ground research (e.g. McKinley et al., 2017; Nault et al., 2023). The first dependent variable was item memory accuracy in Session 2, coded as 1 (correct) or 0 (incorrect). The fixed effects were sum-coded group (sleep: -0.5, wake: 0.5) and role (director: -0.5, matcher: 0.5).

Random effects were participant, item and list (1/2). The sum score of Session 1 recall accuracy was included as a covariate, which controlled for baseline individual differences, as used in the sleep literature (Berres & Erdfelder, 2021). Common ground formation in the referential communication task, reflected by the change in description length from Trial 1 to 3 in relation to the total number of words used in Trial 1 and 3 ( ˝Yρطي˝Yσ ℵ˝Yρط˝Yσ , was calculated separately for each participant, centred and then included as a covariate. Buildmer simulations were used to identify the maximal random effects structure. Results revealed significant effects of group ($\beta$ = -0.32, *SE* = 0.16, *z* = -1.97, *p* = .049), role ($\beta$ = -0.92, *SE* = 0.15, *z* = -6.00, *p* < .001), and Session 1 recall accuracy ($\beta$ = 4.25, *SE* = 0.84, *z* = 5.03, *p* < .001). As shown in Figure 3, accuracy of recognising old items was higher in the sleep (*vs.* wake) group, for items encoded in the director (*vs.* matcher) role, and when Session 1 recall accuracy was high. The interaction between group and role was not significant (*p* = .530).

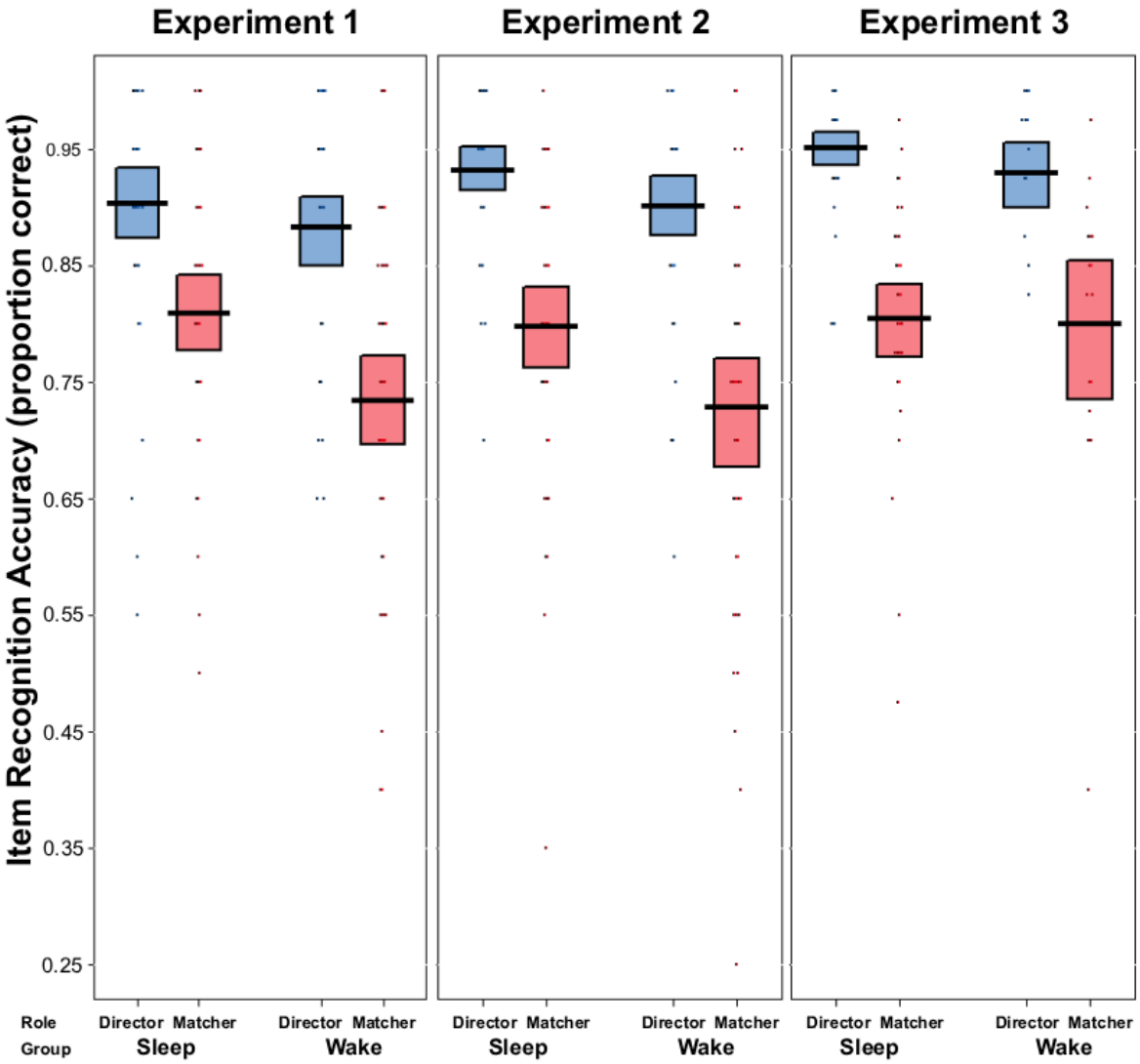


**Figure 3.** Item memory accuracy (proportion of correctly recognised items; 1 = all items correctly recognised) by group and role across Experiments 1-3. Note: Behavioural results for Experiment 3 include only N=39 in the Sleep group and N=19 in the Wake group given its focus on the neural predictors of sleep effects.

Source memory, remembering with whom a particular object was discussed with, was analysed with a comparable logit mixed-effect model. The dependent variable was the accuracy of source memory in Session 2, coded as 1 (correct) or 0 (incorrect). The fixed effects, covariates and random effects structure determined from Buildmer simulations were

identical to those described for the analysis of item memory. Results revealed significant effects of group ($\beta$ = -0.22, $SE$ = 0.11, $z$ = -2.04, $p$ = .041) and Session 1 source accuracy ($\beta$ = 2.83, $SE$ = 0.39, $z$ = 7.21, $p$ < .001). As shown in Figure 4, accuracy of recalling the source of old items was higher in the sleep (*vs.* wake) group and when Session 1 source accuracy was high. None of the remaining effects or interactions reached significance ($p$s > .528).

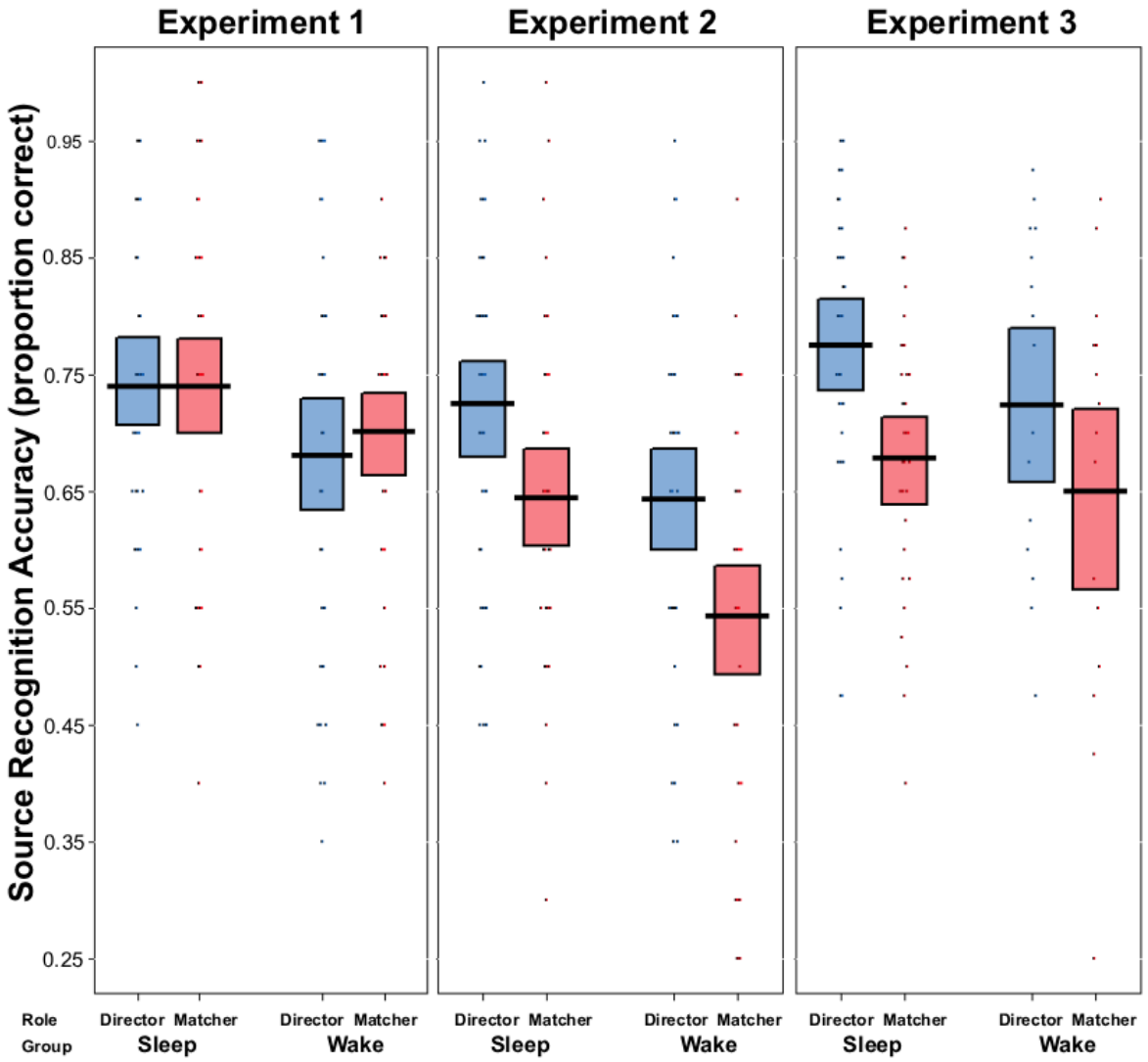


**Figure 4.** Source memory accuracy (proportion of correctly recalled source; 1 = all sources correctly identified) by group and role across Experiments 1-3. Note: Behavioural results for Experiment 3 include only N=39 in the Sleep group and N=19 in the Wake group given its focus on the neural predictors of sleep effects.

As suggested by an anonymous reviewer, we supplemented our pre-registered mixed-effect modelling analysis with exploratory multinomial processing tree (MPT) models. MPT models are a class of cognitive models that quantify and delineate distinct cognitive processes underlying responding in a behavioural task. For a task measuring item and source memory,

as in the current study, Bayen et al. (1996) developed the two-high threshold model of source monitoring to quantify item (*D*: The probability of recognising an old/new item) and source memory capacity (*d*: Having recognised an old item, the conditional probability of recalling its source) delineated from guessing and other non-memory processes.

MPT modelling supplemented our mixed-effects models by providing additional insight into group differences in memory after separating out guessing/non-memory processes. However, our core conclusions remain derived from the mixed modelling analyses for the following reasons. First, mixed-effect modelling was the pre-registered analysis to investigate our hypotheses as this technique has been used in related memory and common ground research (McKinley et al., 2017; Nault et al., 2023). Second, and relatedly, our sample size was based on the expected group effect on item/source memory responses analysed in a mixed-effect model. MPT models were therefore not considered in our power analyses which require additional a-prior assumptions about *every* estimated parameter (e.g., guessing) in the model (Schnuerch et al., 2020). Therefore, it is unclear if our sample size lends itself to an appropriately-powered MPT model, warranting some caution to their findings on the basis of statistical power.

We configured Bayen et al.'s (1996) source monitoring model to analyse the memory data in Session 2, using a partial-pooling approach in a Bayesian framework (Heck et al., 2018). Separate models were built for each group to compare group-level memory parameters across groups, then we sampled the mean parameter difference and assumed a statistically reliable effect if the Bayesian credibility interval did not pass zero. For the sake of brevity, we only present the key findings from these analyses here and direct the reader to the Supplementary Materials for a full overview of the model pathways and parameters. In contrast to the mixed-effect modelling analysis, MPT analysis revealed no reliable group

differences in item (Sleep group D = .80 [.77 .84], Wake group D = .77 [.74 .79], ΔD = .04 [-.01 .08]) or source memory (Sleep group d = .61 [.53 .69], Wake group d = .48 [.38 .58], Δd = .13 [-.004 .25]).

*Common ground production task*

The common ground production task in Session 2 indicated whether participants maintained common ground ~12 hours after encoding. Two analyses were conducted on these data. The first compared the length of descriptions between old and new images and whether this differed across groups. We refer to this as item-specific conversation memory because it captures memory for shared referents overall, without considering whether the conversation partner was the same or different between sessions. The second considered only old items and compared the length of descriptions across groups and roles when the item was encoded/probed with a matching *versus* mismatching source. We refer to this as partner-specific common ground because it reflects knowledge tied to the shared conversational history with a specific individual.

First, a linear mixed-effect model was used to analyse the length of descriptions produced for old and new items, with fixed effects of group (sleep: -0.5, wake: 0.5) and condition (new: -0.5, old: 0.5). Session 1 item-specific conversation memory (i.e., the normalised difference in total word count between Trial 1 and 3 of the referential communication task ″Yρطي″Yσ ℐ″Yρطى″Yσ) was centred and included as a covariate. Buildmer simulations revealed the maximal random effects structure for participants, items, and list. Analysis revealed a significant effect of condition ($\beta$ = -0.35, *SE* = 0.06, *t* = -5.94, *p* < .001) and a significant interaction between group and condition ($\beta$ = 0.35, *SE* = 0.12, *t* = 2.95, *p* = .003). As shown in Figure 5, participants used fewer words to describe old compared to new items, and this difference between old *vs*. new items was larger in the sleep

group ($\beta = 0.53$, $SE = 0.08$, $t = 6.38$, $p < .001$) than the wake group ($\beta = 0.18$, $SE = 0.09$, $t = 2.09$, $p = .037$). The effect of group was not significant ($p = .215$).

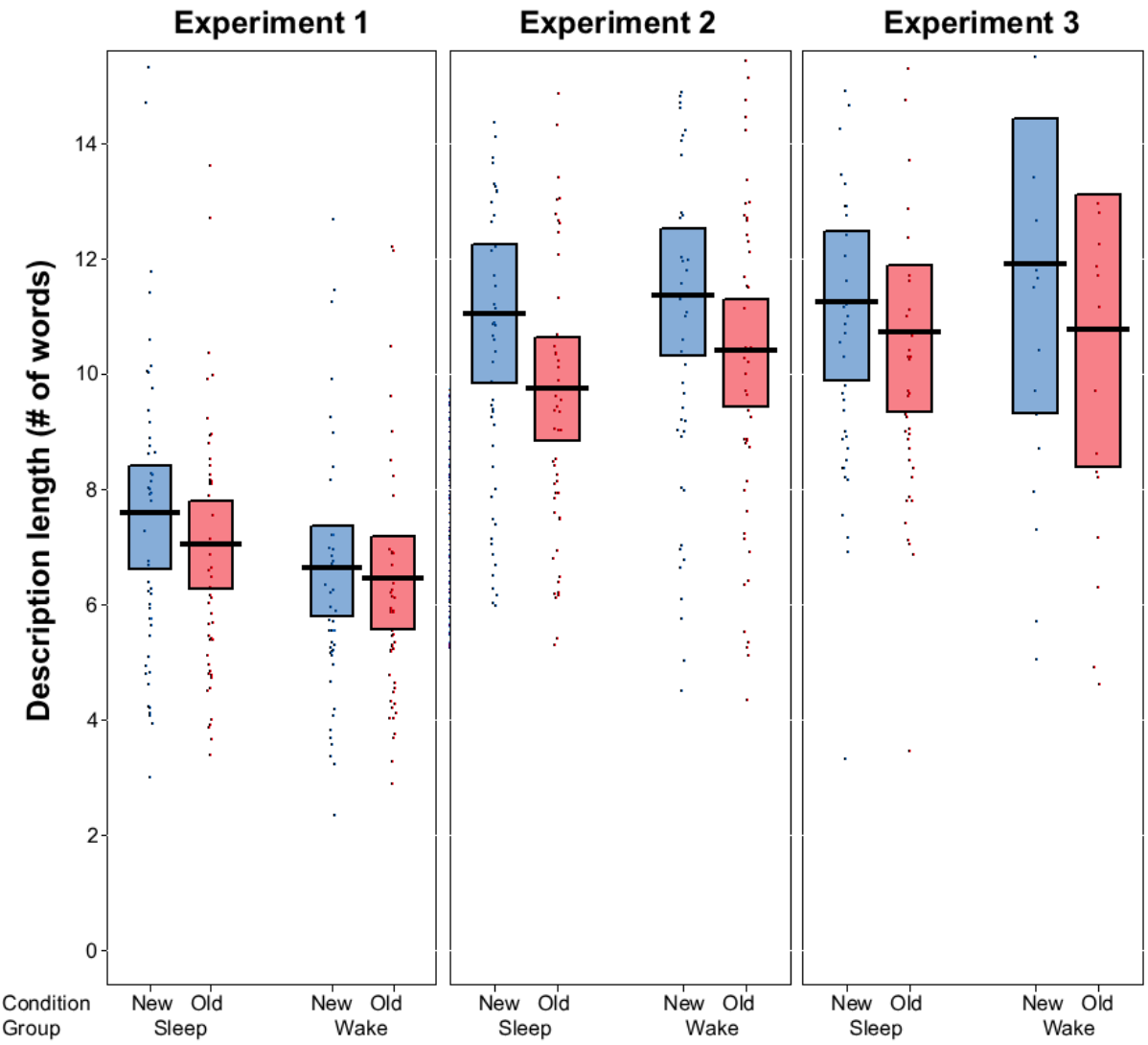


**Figure 5.** Item-specific conversation memory: the average number of words used to describe items in the common ground production task by group and condition across Experiments 1-3. Note 1: The common ground production task in Experiment 1 was computerised and participants typed descriptions item-by-item, whereas common ground production in Experiments 2 and 3 was assessed in face-to-face interactions with participants giving verbal descriptions for each item in a 5 × 4 grid (meaning they included lexical disfluencies and location information). Note 2: Behavioural results for Experiment 3 include only N=39 in the Sleep group and N=19 in the Wake group given its focus on the neural predictors of sleep effects.

Second, a linear mixed-effect model was used to analyse the length of descriptions produced for old items in this task, with fixed effects of group (sleep: -0.5, wake: 0.5), role (director: -0.5, matcher: 0.5), and source match (match: -0.5, mismatch: 0.5). Session 1 common ground (i.e., the normalised difference in total word count between Trial 1 and 3: ˝Ypطي˝Yσ ⁒ ˝Ypط˝Yσ ) was centred and included as a covariate, and Buildmer simulations were used to identify the maximal random effects structure. Results showed no significant effects or interactions (all *p*s > .094). A summary of results can be found in Figure 6.

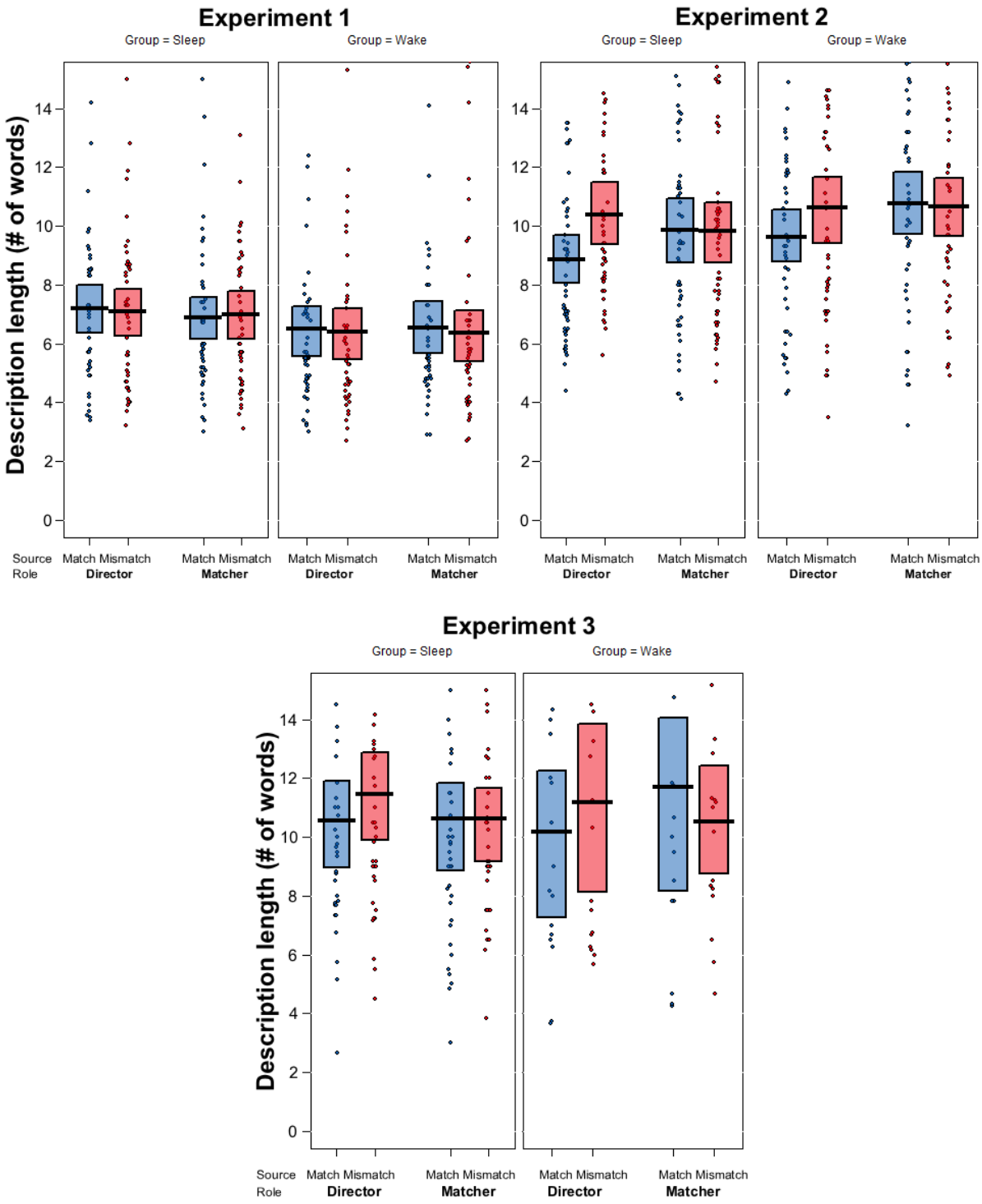


**Figure 6**. Partner-specific common ground: the average number of words used to describe old items in the common ground production task by group, role and source match across Experiments 1-3. Note: Behavioural results for Experiment 3 include only N=39 in the Sleep group and N=19 in the Wake group given its focus on the neural predictors of sleep effects.

**Discussion**

The results from our pre-registered mixed-effect modelling analyses provided evidence that several aspects of conversational memory were enhanced in the Sleep relative to the Wake group. Participants who slept were more likely to recognise old items after 12 hours, suggesting a better memory of *what* was discussed in conversation, and were also more accurate in recalling source information (i.e., which partner an item had been discussed with). These results align with the broader literature showing that sleep strengthens episodic memory for content and social context. Numerically speaking, the Sleep group also displayed enhanced memory in the exploratory MPT models, however this did not reach reliability on either item or source memory.

After 12 hours, the use of item-specific conversation memory (the relative reduction in description length for old compared to new items) was larger in the Sleep than the Wake group, consistent with overnight sleep having an enhancing impact on memory. However, we did not observe a significant effect of partner-specific common ground; description lengths did not differ when items were described for the original *versus* a new partner. This contrasts with prior work showing that speakers typically elaborate more with new partners who lack a shared history (Horton & Gerrig, 2002; Isaacs & Clark, 1987; Schober & Clark, 1989; Yoon & Brown-Schmidt, 2014)**.** One likely reason for this is that our paradigm provided only minimal cues to partner identity in the common ground production task (i.e., names on a screen) and no face-to-face interaction, conditions under which partner-specific common ground is often attenuated (Brown-Schmidt, 2009; Brown-Schmidt et al., 2015).

The absence of partner-specific effects on common ground raises questions about the ecological validity of this paradigm, as its non-interactive elements may have hindered encoding and retrieval of partner-specific representations. To address this, Experiment 2 introduced live, in-person communication during both the referential communication and

common ground production tasks. In addition, analysis of Experiment 1 PSQI responses revealed that the average reported bedtime was 1:10 AM and the average getting up time was 9:40 AM. Therefore, we adjusted the session start times for Experiment 2 so that the wake group completed Session 1 at 9 AM and Session 2 at 9 PM the same day and the sleep group completed Session 1 at 9 PM and Session 2 at 9 AM the next day.

*Experiment 2*

**Methods**

*Participants*

Our pre-registered target sample size was identical to Experiment 1 (96 participants). To achieve this, a total of 114 participants were recruited from the University of [anonymised for peer-review]. Sixteen participants were excluded because they or one of their partners failed to return for Session 2. The final sample was ninety-eight (86 females and 12 males; $M_{age}$ = 19.08 years; *SD* = 1.25; 48 in Wake group, and 50 in Sleep group), who took part in the study across 32 groups. Eligibility criteria, consent and payment was the same as in Experiment 1.

*Sleep measures*

The PSQI, ESS, rMEQ and SSS were administered as detailed in Experiment 1.

*Referential communication task*

This task and image stimuli were identical to those described in Experiment 1, except that participants completed the referential communication task live, in a face-to-face context rather than on adjacent computers. The director and matcher sat on opposite sides of a table, with a screen divider placed in the middle to prevent participants seeing each other's grid and images (see Figure 2). The director described the arrangement of images from printed and

laminated grids, and the matcher followed the director's instructions to physically arrange their set of images (20 on each trial, individually printed and laminated) into a printed and laminated blank grid.

*Memory task*

The procedure for this task was identical to Experiment 1.

*Common ground production task*

Common ground production was measured in Session 2 by a live, face-to-face referential communication task, with the same instructions as in the referential communication task completed in Session 1 (i.e., to either describe images to the other person (as a director) or to identify described images and place them in an empty grid (matcher)). Across eight blocks, participants worked as the director four times and as the matcher four times, of which half were with partner B and half with partner C (i.e., two blocks in each role/partner combination). Each block consisted of one grid with 20 images, and all 160 images were encountered across the eight blocks. Half of the images were old and half were new. Of the old images, 40 were seen in the same role as in the Session 1 referential communication task and 40 were seen in a different role; half of each was seen with the same partner as in the Session 1 referential communication task (source match) and half was seen with a different partner (source mismatch).

The number of words used to describe each image was calculated following the same process of transcription and coding as in Session 1.

*Procedure*

The sequence of tasks was the same as that described in Experiment 1.

## Results

### *Sleep measures and manipulation check*

Over the last month, the average sleep time reported by participants in the PSQI was 0:48 AM, and the average wake time was 9:01 AM. Linear mixed-effects models showed no significant differences between groups on any sleep measure (PSQI: $\beta = 0.32$, $SE = 0.77$, $t = 0.42$, $p = .679$; ESS: $\beta = 0.36$, $SE = 0.98$, $t = 0.37$, $p = .713$; rMEQ: $\beta = 1.19$, $SE = 0.72$, $t = 1.66$, $p = .101$; SSS, Session 1: $\beta = 0.32$, $SE = 0.22$, $t = 1.48$, $p = .144$; SSS, Session 2: $\beta = -0.04$, $SE = 0.31$, $t = -0.13$, $p = .895$).

### *Referential communication task*

A linear mixed-effect model (following the same structure detailed in Experiment 1) showed a negative effect of trial centred, indicating a linear decrease across the three trials ($\beta = -2.30$, $SE = 0.03$, $t = -88.87$, $p < .001$). Thus, replicating Experiment 1, participants used significantly fewer words to describe the same image from Trial 1 to 3 (see Table 1).

### *Memory task*

A logit mixed-effect model was used to analyse item memory accuracy (see Experiment 1 for model structure). This showed significant effects of group ($\beta = -0.53$, $SE = 0.22$, $z = -2.39$, $p = .017$), role ($\beta = -1.08$, $SE = 0.18$, $z = -6.14$, $p < .001$) and Session 1 item accuracy ($\beta = 3.56$, $SE = 0.73$, $z = 4.87$, $p < .001$). Replicating Experiment 1, accuracy was higher in the sleep *vs.* wake group, when they encoded the item as a director *vs.* matcher, and when Session 1 item accuracy was high. The interaction between group and role was not significant ($p = .990$, see Figure 3).

Analysis of source memory accuracy (see Experiment 1 for model structure) revealed significant effects of group ($\beta$ = -0.45, *SE* = 0.14, *z* = -3.27, *p* = .001), role ($\beta$ = -0.19, *SE* = 0.09, *z* = -2.15, *p* = .032) and Session 1 source accuracy ($\beta$ = 2.54, *SE* = 0.38, *z* = 6.67, *p* < .001). Participants were more likely to recall which partner they saw an item with when they slept between sessions 1 and 2, when they had encoded that item as the director, and when Session 1 source memory accuracy was high. The interaction between group and role was not significant (*p* = .879, see Figure 4).

As in Experiment 1, exploratory MPT modelling analyses revealed no reliable group differences in item memory (Sleep group *D* = .78 [.73 .83], Wake group *D* = .77 [.73 .81], Δ*D* = .01 [-.05 .07]). However, consistent with the mixed-effect modelling analysis, there was a reliable group difference in source memory (Δ*d* = .19 [.06 .32]), with enhanced source memory in the Sleep (*d* = .66 [.56 .76]) relative to the Wake group (*d* = .47 [.39 .56]; see Supplementary materials for full results).

*Common ground production task*

The number of words that participants used to describe each image was analysed using two linear mixed-effect models, as described in Experiment 1. Analysis of descriptions for old *versus* new items revealed a significant effect of condition ($\beta$ = -1.18, *SE* = 0.11, *t* = -10.53, *p* < .001) and a significant interaction between group and condition ($\beta$ = 0.53, *SE* = 0.22, *t* = 2.38, *p* = .018). Consistent with Experiment 1, participants used more words to describe a new *vs.* old item, and this difference was larger in the sleep group ($\beta$ = 1.44, *SE* = 0.16, *t* = 9.22, *p* < .001) than the wake group ($\beta$ = 0.91, *SE* = 0.16, *t* = 5.70, *p* < .001). The effect of group was not significant (*p* = .432, see Figure 5).

Analysis of group, role and source match on the production of common ground for old items showed no significant effects or interactions (all *p*s > .086).

**Discussion**

Consistent with Experiment 1, our pre-registered mixed-effect modelling analysis found that both item and source memory were enhanced after overnight sleep compared to daytime wakefulness, as was the use of item-specific conversation memory (the MPT modelling results also revealed enhanced source memory in the Sleep group). Participants produced shorter descriptions for old than new items, indicating that previously established conversation memory facilitated more efficient referring expressions. This old-new reduction was larger in the Sleep group than in the Wake group, suggesting that sleep strengthened the accessibility and use of item-specific knowledge, maintaining efficient communication strategies over a 12-hour delay. Experiment 2 did not show any evidence that participants encoded a partner-specific representation of common ground in memory, and this was not modulated by sleep.

Taken together, Experiments 1 and 2 provide converging evidence that conversational memory and pragmatic efficiency is enhanced after a period of overnight sleep. Overnight sleep strengthened item and source memory, with stronger and more consistent sleep effects emerging on source memory. In addition, overnight sleep maintained more efficient use of common ground, specifically enhancing the accessibility of previously shared (old *vs.* new) information, over a 12-hour interval. These findings suggest that sleep benefits not only the retention of conversational content but also the communicative deployment of that knowledge in social interaction.

Nevertheless, the mechanisms underlying these sleep-related effects remain to be determined. Sleep could act passively, by protecting memories from interference (Yonelinas et al., 2019), or actively, via offline consolidation during non-REM sleep (Born & Wilhelm, 2012; Klinzing et al., 2019; Rasch & Born, 2013). Experiment 3 therefore examined

oscillatory activity during NREM sleep to test whether neural markers of active consolidation predict the retention of conversational memory and common ground.

*Experiment 3*

Sleep-based memory reactivation is hypothesised to facilitate the transfer of newly encoded, labile episodic memories into stable long-term representations. This process involves the redistribution of memory traces from the hippocampus to neocortical networks that support long-term knowledge (Born & Wilhelm, 2012; Denis & Cairney, 2023). Growing evidence suggests that such consolidation is supported by the finely tuned interplay of distinct neuronal oscillations during non-REM (NREM) sleep, including cortical slow oscillations (~1 Hz), sleep spindles (~11 - 16 Hz) and hippocampal sharp-wave ripples (80 - 150 Hz). Sharp-wave ripples reflect transient reactivation events that are nested in the troughs of sleep spindles, which are thought to reinstate memory-related neural activity (Antony et al., 2018) and promote synaptic plasticity (Fernandez & Lüthi, 2020). In turn, spindles preferentially couple with the depolarising up-phase/peak of slow oscillations (SO), corresponding to periods of heightened cortical excitability (Neske, 2016). This precise coordination is thought to enable the integration of hippocampal reactivations into long-term, cortical memory traces.

Consistent with this account, polysomnography studies have reported positive associations between spindle activity and subsequent memory retention (Denis et al., 2021; Gais et al., 2002; Schabus et al., 2004; Schmidig et al., 2024). Moreover, the coupling between spindles and slow-oscillations (SO-spindle coupling) appears particularly important for memory consolidation (Ng et al., 2025). Specifically, the degree to which spindle peaks are aligned with the SO up-phase, as well as the overall number of SO-spindle coupling events, both correlate positively with post-sleep memory performance (Denis et al., 2021; Halonen et al., 2021; Helfrich et al., 2018; Mikutta et al., 2019; Muehlroth et al., 2019; Ng et

al., 2025). Together, these findings support the view that sleep-related memory benefits are at least partially sourced by active neural processes that coordinate hippocampal reactivation and neocortical integration during NREM sleep. Experiment 3 directly tested this hypothesis, examining whether oscillatory markers of sleep-based memory consolidation predict retention of conversational memories and common ground.

Experiment 3 used near-identical procedures to Experiment 2. The key difference was that the delay between sessions 1 and 2 was reduced from 12 hours to ~4 hours, and participants in the sleep group took a ~2 hour long, polysomnographically-recorded nap in the lab in between sessions (see Figure 1; participants in the wake group spent an afternoon awake). Because this experiment was primarily focused on uncovering possible correlations between sleep electrophysiology and memory/common ground retention, we prioritised participant recruitment in favour of the Sleep group to satisfy statistical power for such correlational effects (see the *Participants* section for details of our sample size). Therefore, the overall sample size of Experiment 3 (N = 58) was not only smaller than the previous two experiments, but was also imbalanced in favour of the Sleep group (Sleep N = 39, Wake N = 19). Between-group comparisons of behavioural responses were therefore viewed as exploratory and are reported in the Supplementary Materials.

The Sleep group's EEG data was used to extract measures of oscillatory activity during NREM sleep that have been linked to sleep-related memory consolidation processes (Denis & Cairney, 2023). If sleep is actively involved in consolidating memories for conversation and common ground, then the retention of memory/common ground should correlate with some of these measures. The specific measures of oscillatory activity were sleep spindle density (number of spindles per minute), the percentage of spindles coupled to slow oscillations, and the preferred phase of slow oscillations to which spindles preferentially coupled. We define retention here as the difference in performance from Session 1 to Session

2. To enable a comparable retention measure for common ground in Experiment 3, we therefore introduced a common ground production task in Session 1. The experiment was pre-registered ahead of data collection with the following hypotheses:

1. Replicating previous findings (McKinley et al., 2017; Nault et al., 2023; Experiments 1 and 2 above), we predicted that pairs of interlocutors would develop common ground in the referential communication task, indicated by a significant reduction in the number of words used to describe a given image from the first to the third time it is discussed.
2. Spindle density (number of spindles/minute) would be positively correlated with retention of item and source memory and the retention of item-specific conversation memory and partner-specific common ground.
3. The percentage of coupled sleep spindles (spindles that occur within a detected slow oscillation, out of all detected sleep spindles) would be positively correlated with the retention of item and source memory and the retention of item-specific conversation memory and partner-specific common ground.
4. The extent to which spindles couple towards the up-phase/peak of the SO would be positively correlated to the retention of item and source memory and retention of item-specific conversation memory and partner-specific common ground (Halonen et al., 2021; Helfrich et al., 2018; Mikutta et al., 2019; Muehlroth et al., 2019; Ng et al., 2025).

## Methods

### *Participants*

Sixty participants were recruited in total from a participation scheme at the University of [anonymised for peer-review]. Forty-one of these participated in the sleep group. The data

from two participants in the sleep group were excluded from analyses, one for failing to achieve >10 minutes of NREM sleep during the 2-hour nap (pre-registered criteria), and a second due to data loss in the memory tasks. This left a final sample of 39 in the sleep group who contributed data to the analyses (30 females and 9 males; $M_{age}$ = 20.03 years; *SD* = 2.53). Whilst our primary focus related to behaviour and EEG responses in this sleep group, laboratory and time restrictions meant that only two participants (out of the four participating in a particular session) could take part in the sleep group. Therefore, we classified these remaining participants as wake participants because they went about their daily activities in between sessions while the sleep participants slept in the laboratory. Nineteen participants took part in the wake group (18 females and 1 male; $M_{age}$ = 19.74 years; *SD* = 2.58). Eligibility criteria were the same as in Experiments 1 and 2.

Pre-registration stated that data collection would stop after reaching 40 usable data sets/participants in the sleep group. This was based on an a-priori power analysis performed in GPower (version 3.1), which calculated the minimum number of participants required to detect a correlation of $r$ = .40 with 80% statistical power, which revealed the necessity for 37 participants in the sleep group. We based our estimate on a correlational effect since the majority of our hypotheses concern relationships between memory and oscillatory activity during sleep. We selected a correlation of $r$ = .40 as similar sleep EEG studies with comparable measures (e.g., spindle density) have reported similar correlations (e.g., Denis et al., 2021; Gais et al., 2002). Our pre-registration increased the target sample to 40 to make it a multiple of four, since our design permitted a maximum of four participants in a single session.

*Materials and procedure*

Upon arrival to the lab, participants were informed of their group allocation[2]. All participants provided consent before completing the Qualtrics survey assessing sleep habits and behaviours (PSQI, ESS, rMEQ). Participants in the sleep group were then set up with EEG (see *EEG acquisition*). Once EEG set up was complete, all participants completed the SSS for the first time.

Session 1 began with the referential communication task at around 12 PM, adopting the same face-to-face procedures as Experiment 2. Each participant encountered 80 different items, and discussed these with one of two partners, either in the role of director or matcher. Participants then completed the memory task on an individual computer, before completing the first common ground production task (again, with the same procedures as Experiment 2). Participants completed four blocks of this task, involving 40 old and 40 new items.

Session 1 was completed by around 1 PM. At this point, participants in the Sleep group got into bed and were left to sleep for ~2 hours. Participants in the Wake group were asked to go about their daily activities before returning to the lab around 2.5 hours later to complete Session 2. Session 2 began around 4 PM, approximately 45 minutes after participants in the sleep group had woken up to accommodate removal of the EEG electrodes and to alleviate the effects of sleep inertia. All participants completed the SSS for a second time, before completing the second memory task and the second common ground production task. Each task involved the 40 old and 40 new images that were not tested in the corresponding tasks in Session 1). Thus, the additional common ground production task

[2] Participants in the sleep group were asked to arrive 1 hour earlier than wake group participants (this provided extra time for setting up the EEG).

introduced in Experiment 3 involved a different subset of items, meaning that it could not directly influence the depth of encoding for the images tested in Session 2.

*EEG acquisition and pre-processing*

Sleep EEG was measured and recorded using a Natus Embla NDx recording system with RemLogic 4.0 recording software. Scalp EEG electrodes were positioned according to the 10-20 system over frontal (F3 and F4), central (C3 and C4), parietal (P3 and P4), occipital (O1 andO2), and mastoid (M1 and M2) sites, referenced online to Cz with a ground electrode positioned on the forehead, and a sampling rate of 256 Hz. Bilateral electrooculography electrodes were positioned ~1 cm from each eye, and electromyography electrodes were positioned on the thementalis and submentalis (bilaterally). Electrode impedances were kept <5 Ω.

Offline, EEG electrodes were band-pass filtered at 0.5 - 35 Hz, downsampled to 200 Hz, and re-referenced to the contralateral mastoid. Sleep scoring took place in the *CountingSheepPSG* toolbox in Matlab (Ray et al., 2024) and followed standard criteria (Iber et al., 2007). Sleep stages were scored in 30 seconds epochs, and epochs containing body movements/muscle artefacts were tagged and removed from subsequent analyses.

*Sleep spindle detection*

Analysis of sleep spindle and slow oscillation detection/coupling included only epochs scored as N2 or N3/SWS (NREM sleep). Sleep spindles were automatically detected using a wavelet-based decomposition method (Denis et al., 2021). First, for each participant, we determined the peak spindle frequency as the most prominent frequency peak between 11 - 16 Hz across all electrodes. The raw EEG signal was then subjected to a time-frequency decomposition using a wavelet-based detector, with the peak frequency determined as the

individual's spindle peak (defined above) with a band-width of 3 Hz. Spindle detection was performed on the squared wavelet coefficients after being smoothed with a 100 ms moving average. A spindle was detected whenever the wavelet signal exceeds a threshold of six times the median signal amplitude of NREM sleep for at least 400 ms, therefore accounting for individual variability in spindle amplitude.

*Slow oscillation (SO) detection*

First, the EEG data was bandpass filtered between 0.5 - 4Hz and all positive-to-negative zero crossings identified. Candidate SOs were tagged if two such consecutive zero crossings fell 0.8 - 2 seconds apart, corresponding to 0.5 - 1.25 Hz. SOs in the top quartile of peak-to-peak amplitude were retained (i.e., those with the highest amplitudes).

*SO-spindle coupling*

To identify SO-spindle coupling events, the EEG data was bandpass filtered in each participant's peak spindle frequency range. Then, a Hilbert transform was applied to extract the instantaneous phase of the delta (0.5 - 4Hz) filtered signal and the instantaneous amplitude of the spindle filtered signal. For each detected spindle, the peak amplitude of the spindle was determined. It was then determined whether the spindle peak occurred within the time course (i.e., between two consecutive positive-to-negative zero crossings) of any detected SO. If the spindle peak was found to occur during a SO, the phase angle of the SO at the peak of the spindle was determined. Circular-linear correlations involving phase angle were carried out using radians. To ease interpretation, we present the data in degrees, with a phase angle of 0° corresponding to the peak of the SO.

## Results

### *Referential communication task*

The average reported sleep onset time in the month before the experiment was 00:40 AM, with an average wake up time 08:09 AM. Before investigating the relationship between sleep oscillatory activity and behaviour, we first verified that common ground was established in the Session 1 referential communication task. This was achieved using a linear mixed-effect model, as in Experiments 1 and 2. The results showed a negative effect of trial centred, suggesting a linear decrease across the trials ($\beta$ = -2.60, $SE$ = 0.07, $t$ = -36.23, $p < .001$). Hence, participants used fewer words to describe each image from trial 1 to trial 3, indexing the formation of common ground (see Table 1).

### *EEG results*

Macro-level sleep physiology is summarised in Table 2. On average, participant's total sleep time was around 90 minutes.

**Table 2.** Mean time and percentage spent in different sleep stages. Standard deviations are presented in parentheses.

| Wake time (min) | N1 time (min) | N2 time (min) | SWS time (min) | REM time (min) | Total sleep time (min) | N1 % | N2 % | SWS % | REM % |
|---|---|---|---|---|---|---|---|---|---|
| 31.74 (24.52) | 13.15 (9.25) | 51.17 (15.60) | 10.06 (14.09) | 12.60 (11.96) | 86.99 (24.19) | 15.60 (9.95) | 60.71 (16.11) | 10.78 (15.15) | 12.92 (11.60) |

*Note: N1 = Stage 1 sleep; N2 = Stage 2 sleep;* SWS = Slow-wave sleep; REM = Rapid eye movement.

Concerning SO-spindle coupling, a Rayleigh's test of non-uniformity revealed that the distribution of spindle peaks around SOs was significantly non-uniform (adjusted alpha level of $p = .0063$) across participants at all eight electrodes ($z$'s > 7.56; $p$'s < .000382). Although we detected coupling events over the whole scalp, our analyses involving SO-spindle coupling focused on averaged data across the two frontal electrodes (F3 and F4), based on recent evidence highlighting a particular affinity between frontal coupling events and memory (Ng et al., 2025). This analysis revealed that most participants displayed significant non-uniformity (32/39; 82%), with a mean coupling phase of 25.94° (+/- 50.79°). This corresponds to just past the positive peak of the SO (see Figure 7) and is consistent with prior work indicating a preference for spindles to couple to this part of the SO (Denis et al., 2021, 2025; Muehlroth et al., 2019; Ng et al., 2025). See Table S6 in the Supplementary Materials for descriptive statistics on micro-level sleep physiology.

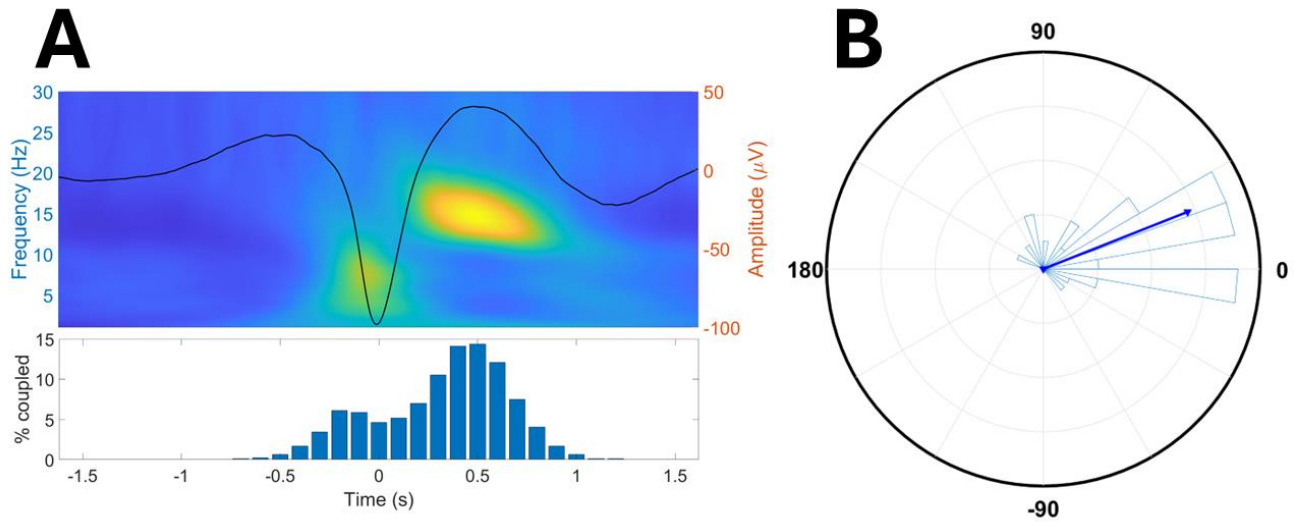


**Figure 7. A**) Top) The time frequency representation of all SOs coupled to a sleep spindle (-1500 - 1500 ms, centred on the trough of the SO), averaged to the two frontal electrodes. The averaged time-domain SO is superimposed. Bottom) A peri-event histogram displaying the distribution of coupled spindles relative to the average SO, in 100 ms time bins. The

combined sum of percentages across all bars is equal to100%. **B)** A circular phase plot displaying the proportional distribution of participant spindle phase coupling, averaged to the two frontal electrodes. The group mean is represented by the blue arrow. A phase angle of 0° corresponds to the peak of the SO.

Three pre-registered measures of NREM sleep-related oscillatory activity were tested for correlations with behaviour: 1) sleep spindle density (number of spindles per minute), averaged across the whole scalp; 2) the percentage of spindles coupled to slow oscillations, and 3) the preferred phase of slow oscillations to which spindles coupled. The latter two measures were averaged over frontal electrodes (F3, F4). These measures were correlated with four behavioural measures of memory/common ground retention, calculated for each participant: item memory, source memory, item-specific conversation memory, and partner-specific common ground. These retention measures aimed to capture the relative change in memory and common ground across the sleep period, which are often more sensitive in revealing correlations with sleep physiology than absolute memory performance(e.g., Schabus et al., 2004). Item and source memory were quantified by subtracting item/source memory accuracy in Session 1 from Session 2, with positive values indicating superior performance in Session 2. The retention of item-specific conversation memory was calculated by first calculating the difference in description length between new and old items (new - old items) within each session and then subtracting this difference score in Session 1 from Session 2 (i.e., (Session $2_{new}$ – Session $2_{old}$) - (Session $1_{new}$ – Session $1_{old}$)). Positive values indicate that the relative use of item-specific conversation memory increased from Session 1 to Session 2. The retention of partner-specific common ground was examined separately for old items, by first calculating the difference in description length between source mismatch and source match trials (source mismatch – source match), and then subtracting the Session 1

difference from the Session 2 difference (i.e., (Session $2_{mismatch}$ – Session $2_{match}$) - (Session $1_{mismatch}$ – Session $1_{match}$)). Positive values indicate greater retention or use of partner-specific common ground in Session 2. Table 3 presents mean values of these different measures in each session. See Supplementary Materials for exploratory analyses of memory and common ground, akin to Experiments 1 and 2.

**Table 3.** Mean values for each behavioural measure in Experiment 3 (standard deviations in parentheses). For the correlational analyses, values in Session 1 (S1) were subtracted from the corresponding value in Session 2 (S2).

| Memory (Proportion correct) | | | | Word count measures | | | |
|---|---|---|---|---|---|---|---|
| Item memory | | Source memory | | Item-specific conversation memory (new – old) | | Partner-specific common ground (old source mismatch – old source match) | |
| S1 | S2 | S1 | S2 | S1 | S2 | S1 | S2 |
| 0.89 | 0.87 | 0.77 | 0.69 | 1.81 | 0.54 | 0.43 | 0.39 |
| (0.08) | (0.06) | (0.12) | (0.11) | (1.63) | (1.47) | (1.47) | (2.14) |

Correlations between the four behavioural measures and three measures of sleep oscillatory activity are summarised in Table 4. Correlations involving spindle density and % of coupled spindles utilised Pearson's correlations, whilst correlations involving preferred coupling phase utilised circular-linear correlations using the *circstats* toolbox in Matlab (Berens, 2009). As can be seen, there were no statistically significant correlations (all $p$'s > .106).

**Table 4.** The relationship between measures of memory and common ground retention and sleep oscillatory activity.

| | Spindle density | % of coupled spindles | Preferred coupling phase |
|---|---|---|---|
| Item memory retention | $r = .03$, $p = .870$; $BF_{10} = 0.36$ | $r = .05$, $p = .785$; $BF_{10} = 0.37$ | $r = .23$, $p = .368$ |
| Source memory retention | $r = .11$, $p = .491$; $BF_{10} = 0.44$ | $r = .25$, $p = .120$; $BF_{10} = 1.03$ | $r = .33$, $p = .113$ |
| Retention of item-specific conversation memory | $r = .25$, $p = .120$; $BF_{10} = 1.03$ | $r = -.26$, $p = .106$; $BF_{10} = 1.12$ | $r = .31$, $p = .163$ |
| Retention of partner-specific common ground[3] | $r = .02$, $p = .917$; $BF_{10} = 0.36$ | $r = .11$, $p = .492$; $BF_{10} = 0.44$ | $r = .03$, $p = .984$ |

In light of the statistically non-significant correlations, we calculated Bayes Factors ($BF_{10}$) using the *correlationBF* function from the *BayesFactor* R package (Morey & Rouder, 2026) to determine the degree of evidence in support of the null hypothesis reflecting no correlations. This was not carried out for the circular linear correlations of phase angle however, as to our knowledge there is no framework for obtaining Bayes factor for such

[3] Exploratory analyses tested correlations of sleep oscillatory activity with director partner-specific common ground, since common ground distinguished matched and mismatched sources specifically for directors in Experiment 2, but none of the correlations reached significance (all $p$'s > .310).

correlations. The eight Bayes factors for the remaining correlations are presented in Table 4, and ranged from anecdotal evidence in favour of the null (0.36) and alternative hypothesis (1.12, with 1 representing no evidence - see Lee & Wagenmakers, 2014), although the former concerns a negative correlation which was not the predicted direction. Given that the majority of the Bayes factors were <1, the null hypothesis may be slightly more consistent with the data overall, although this support appears fairly weak at best.

*Coupled spindle density*

In line with the pre-registration, an exploratory analysis examined the relationship between coupled spindle density and behaviour. That is, the number of spindles coupled to a slow oscillation per minute, which has been shown to predict memory retention (Denis et al., 2021, 2025). We built four hierarchical regression models, one for each behavioural measure. In the first step of each model, behaviour was predicted by uncoupled spindle density. In the second step, coupled spindle density was entered as a predictor alongside uncoupled spindle density, and the two steps/models were compared using the *anova* function in R. This allowed us to determine whether coupled spindle density accounts for a significant amount of variance in behaviour, after controlling for uncoupled spindle density (Denis et al., 2021). Beta coefficients and p-values for the second steps are shown in Table 5.

**Table 5.** The relationship between coupled and uncoupled spindle density on behavioural measures. Beta coefficients and p-values are taken from the second of model building.

| | Coupled spindle density | Uncoupled spindle density |
|---|---|---|
| Item memory retention | $B = 0.01$, $p = .644$ | $B =- 0.00$, $p = .716$ |
| Source memory retention | $B = 0.04$, $p = .167$ | $B = -0.01$, $p = .710$ |

| | | |
|---|---|---|
| Item-specific conversation memory | $B$ = -0.55, $p$ = .093 | **$B$ = 0.82, $p$ = .014** |
| Partner-specific common ground retention | $B$ = 0.54, $p$ = .320 | $B$ = -0.54, $p$ = .319 |

Coupled spindle density did not account for a significant proportion of variance in the retention of any behavioural measure (all $p$'s > .093). In contrast, uncoupled spindle density was positively correlated with the retention of item-specific conversation memory ($p$ = .014), with more uncoupled spindles predicting a greater reduction in description lengths for old items (relative to new items) after sleep.

**Discussion**

Contrary to predictions, none of the sleep oscillatory indices showed a significant relationship with behavioural retention measures. Follow up Bayesian analyses revealed only anecdotal support for the null hypothesis. Therefore, our confirmatory analyses yielded inconclusive evidence into whether sleep may actively consolidate memory and common ground. However, a pre-registered exploratory analysis revealed that uncoupled spindle density predicted item-specific conversation memory retention. Uncoupled spindles are defined as solitary spindles during NREM sleep, occurring independently from a slow oscillation. Although slow oscillations (and their temporal coupling with sleep spindles) are predominantly believed to facilitate memory consolidation, some findings suggest they can in fact impair memory. For example, Payne et al. (2009) found that SWS duration correlated negatively with veridical recall in a Deese-Roediger-McDermott (DRM) paradigm, with SWS instead argued to support the extraction of gist-like details of a memory. Thus, although our exploratory analyses do not reveal specific SWS impairments on memory, they may

suggest that memory reactivation events occurring independently from slow oscillations (reflected by uncoupled sleep spindles) can still promote some long-term memory benefits.

*General Discussion*

Across three experiments, we examined how sleep influences the maintenance of common ground (i.e. shared referential knowledge built up during conversation) between dyads. Using a referential communication task (McKinley et al., 2017; Nault et al., 2023), we replicated the well-established finding that interlocutors become increasingly efficient over repeated references, as reflected in linear reductions in description length across trials (Clark & Wilkes-Gibbs, 1986; Hupet et al., 1993; McKinley et al., 2017; Nault et al., 2023; Wilkes-Gibbs & Clark, 1992).

We also replicated the generation effect: speakers (Directors) remembered items and their sources better than listeners (Matchers), consistent with enhanced encoding for self-generated material (Jacoby, 1978; Koriat et al., 1991; MacLeod et al., 2010; McKinley et al., 2017; Nault et al., 2023; Riefer et al., 2007; Slamecka & Graf, 1978). Importantly, we extended this literature by showing that the generation benefit applied not only to item memory but also to source memory when communication occurred face-to-face (Experiment 2). This pattern suggests that richer, multimodal social cues (e.g., facial expressions, gestures, prosody) enhance encoding of partner-specific information, thereby strengthening episodic links between content and conversational source.

Importantly, Experiments 1 and 2 revealed that both item and source memory, as well as item-specific conversation memory, were maintained to a greater extent in Sleep group participants who slept overnight compared to Wake group participants who experienced a day awake. These results were obtained in our pre-registered mixed-effect modelling analyses and are consistent with sleep stabilising conversational memory traces, supporting both

the content and efficiency of communication after delay. This finding has parallels to existing work in the cognitive domain of learning (e.g., Berres & Erdfelder, 2021) by showing sleep consolidation effects on information encoded in real-life, interactive social contexts. Supplementary MPT analyses revealed no significant evidence of between group differences in item memory, but in Experiment 2 revealed enhanced source memory in the Sleep group, lending support to the mixed-effects analyses. Speculatively, the more consistent evidence for enhanced source memory in the Sleep group could suggest that sleep may particularly benefit the retention of item-context associations (see also Berres et al., 2025).

Although sleep strengthened the retention of item-specific conversation memory, we did not find any reliable evidence for sleep effects on partner-specific common ground, reflecting shared knowledge or experience between different interlocutors (Clark & Brennan, 1991). This may suggest that the interpersonal component of shared knowledge may be more resistant to consolidation or may require stronger motivational relevance to engage active sleep processes. That said, evidence for partner-specific common ground overall was fairly weak, with only trends for partner-specificity in Experiment 2 in the Director role condition. Thus, future work could attempt to modify the existing paradigm to enhance partner-specific effects. However, the fact that evidence of partner-specificity was stronger in Experiment 2 implies that interactive paradigms may be crucial in supporting partner-specific encoding and/or retrieval (Brown-Schmidt, 2009).

Experiment 3 directly tested whether oscillatory markers of NREM sleep (sleep spindle density, SO-spindle coupling and coupling phase; Denis & Cairney, 2023; Ng et al., 2025) predict the maintenance of conversational memory and common ground. None of these confirmatory measures correlated significantly with behavioural indices of retention. Despite this, follow up Bayes factors only revealed anecdotal support for the null hypothesis (and in some cases the alternative). Additionally, an exploratory analysis revealed a possible role of

uncoupled spindles in maintaining item-specific conversation memory. Therefore, we interpret the findings of Experiment 3 as revealing inconclusive evidence into whether social memories and common ground are actively consolidated during sleep.

There are, however, several methodological and theoretical points to note. Firstly, we may have lacked sufficient statistical power to detect significant associations between EEG activity and behaviour, particularly given trends of small associations in some cases (e.g., $r$ = .20 - $r$ = .30). Although we based our power analysis on detecting a correlation of $r$ = .40, based on prior work using similar EEG measures (e.g., Denis et al., 2021), behavioural sleep effects are typically fairly small and not the most robust (Berres & Erdfelder, 2021; Cordi & Rasch, 2021), so the association between sleep EEG measures and conversational/common ground memories may be smaller than other memory paradigms (e.g., paired associate learning), warranting larger sample sizes. Indeed, the sample sizes for Experiments 1 and 2 were partly based on Mak et al. (2024) who reported that N=48 per group is sufficient for detecting effects of $d > 1$ with 95% power (or 32 participants per group at 80% power). However, a recent meta-analysis suggested that sleep effects on memory may be of smaller magnitude (Berres & Erdfelder, 2021), which suggests that larger samples may be required to detect associations between behavioural sleep effects and sleep physiology in naturalistic memory paradigms. The current study therefore provides an important first estimate of the magnitude of sleep effects in more unconstrained memory contexts, which can inform future power calculations and study design.

Secondly, the between group differences in behaviour (memory and common ground) were non-significant in Experiment 3 (see Supplementary Analysis for analyses with mixed-effect modelling). This contrasts with significant mixed-effect modelling group effects in Experiments 1 and 2, which included larger sample sizes for the between-group comparison and used a longer delay interval (i.e., ~4 hrs vs. ~12 hrs) that included overnight sleep.

Although daytime nap designs are generally considered sufficient to elicit sleep-related memory effects, such effects may be smaller compared to overnight designs (Berres & Erdfelder, 2021; Lo et al., 2014) which are characterised by qualitatively different sleep architecture (Van Schalkwijk et al., 2019). This may mean that naps provide fewer opportunities for memory reactivation events, in turn reducing sleep-related memory enhancement relative to overnight sleep (Lo et al., 2014). The duration of the delay is also relevant when considering wakefulness conditions, since a longer delay affords more opportunity for interference from new memories. This is particularly relevant for the current work since engaging in conversation forms an integral part of everyday life for most adults.

Finally, given that we did not find significant evidence for an active role of sleep, it seems reasonable to consider alternative mechanisms to explain sleep-related memory benefits, including those observed in Experiment 1 and Experiment 2. Rather than actively consolidating memories via memory replay processes (Denis & Cairney, 2023; Klinzing et al., 2019; Rasch & Born, 2013), it has been proposed that sleep may enhance memories by providing a temporary 'protective shield' against interference from new memories; this cannot be achieved to the same extent during wake (Ellenbogen et al., 2006; Yonelinas et al., 2019). This overnight protection from interference may be particularly critical to the consolidation of social memories, since daytime social interactions are more closely related to the initial encoded experience, and are more likely to occur in the wake than the sleep group. Taken together, however, the present findings do not allow us to conclusively adjudicate between active consolidation accounts and more passive, interference-based explanations of sleep-related memory benefits.

**Episodic memory and the maintenance of common ground**

Cognitive models of audience design emphasise that common ground depends partly on episodic memory for prior discourse (Horton, 2007; Horton & Gerrig, 2005). Our data support and extend this framework by showing that the retention of referential knowledge from prior interactions, like other episodic memories, benefits from sleep, regardless of whether the underlying mechanism is active or passive. Drawing on Gaskell's (2024) dual-pathway model of lexical processing, we suggest that recent linguistic experiences are bound together in a memory representation, encoding key conversational elements including the topic and source of conversation (see also Horton, 2007; Horton & Gerrig, 2005). Potentially, this process is supported by the hippocampus whose representations are particularly susceptible to sleep-dependent preservation (Gaskell et al., 2019; Mak et al., 2023).

In our experiments, reencountering an item (i.e., in the memory and common ground production tasks) may have triggered retrieval of its prior conversational context (i.e., in the first referential communication task), resulting in shorter descriptions for old relative to new items. This effect was amplified by overnight sleep, which suggests that sleep preserves episodic traces that underpin efficient reuse of prior referential expressions in discourse. Thus, our findings provide the first evidence that sleep contributes to maintaining pragmatic efficiency in dialogue via memory consolidation processes. It is important to acknowledge that our measure of item-specific conversation memory does not reflect common ground per se, and instead may index increased familiarity with previously encountered images. Although repeated interaction likely supports the formation of shared representations, our design does not allow us to fully dissociate item familiarity from partner-specific common ground. Accordingly, the present findings are more conservatively interpreted as reflecting sleep-related benefits for item-specific conversational memory, rather than for the use of common ground in a strictly partner-specific sense. More generally, this highlights a

challenge in dialogue research, whereby behavioural measures, such as description length, may reflect multiple overlapping processes, including familiarity, abstraction and partner-specific memory.

If the retrieval and efficiency of social memories is greater after sleep relative to wake, why did the EEG markers of reactivation during sleep not predict behavioural outcomes? Alongside the potential methodological and theoretical explanations discussed earlier, another possibility is that any active component of sleep in supporting memory consolidation, driven by memory reactivation processes, is preferentially recruited for information with adaptive or motivational value for the future (Cowan et al., 2021; Diekelmann et al., 2009; Sterpenich et al., 2021; Stickgold & Walker, 2013). The referential labels in our study described familiar objects with pre-existing lexical representations (e.g., "the stripy bird"), which may have reduced the communicative need to actively consolidate these episodes. In other words, actively consolidating a shared memory of a familiar object may be minimally relevant for discussing the image again in the future; active sleep consolidation may be more sensitive to unfamiliar linguistic information (Tamminen et al., 2010, 2013; for a similar discussion see Ball et al., 2025). Future work could test this account by using novel or ambiguous stimuli, such as tangrams (Yoon et al., 2024; Yoon & Brown-Schmidt, 2014), where maintaining idiosyncratic labels would have greater relevance for future interactions.

*Conclusion*

In summary, we have shown that overnight sleep enhances conversational memory and communicative efficiency, consistent with episodic consolidation of discourse-level information. These patterns highlight that sleep not only stabilises factual and semantic memories but also preserves socially-grounded knowledge encoded during a live interaction.

This is a novel finding in the sleep-memory literature, which shows that episodic memory processes (traditionally studied in isolated learning tasks) also contribute to maintaining the coherence and efficiency of dialogue across time. These results bridge cognitive models of memory consolidation with those of pragmatic communication. Sleep may help sustain the representations that enable interlocutors to resume conversation efficiently, thereby linking individual memory mechanisms with the continuity of social interaction. However, the sleep EEG analyses did not conclusively reveal neural correlates that would provide clear support for an active, memory reactivation-based consolidation account. Understanding when and how sleep actively integrates social knowledge into long-term memory will be a critical goal for future research.

*Data Availability Statement*

Datasets generated during and/or analysed during the current study are available in the Open Science Framework repository, https://osf.io/ry4p7/overview?view_only=900b58cfb7354ba68d54d9fb6845de09.

*CRediT Authorship Contribution Statement*

**Y. Q:** Methodology, Software, Formal analysis, Investigation, Resources, Data Curation, Writing – Original Draft, Writing – Review & Editing, Visualization. **L. B:** Methodology, Software, Formal analysis, Investigation, Resources, Data Curation, Writing – Original Draft, Writing – Review & Editing, Visualization. **C. H:** Investigation, Data Curation. **D. D:** Software, Writing – Review & Editing. **G. G:** Conceptualization, Writing – Review & Editing, Supervision, Project administration, Funding acquisition. **H. F:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing – Original Draft, Writing – Review & Editing, Visualization, Supervision, Project administration, Funding acquisition.

## Competing Interests

The authors declare that they have no competing interests.

*References*

Adan, A., & Almirall, H. (1991). Horne & Östberg morningness-eveningness questionnaire: A reduced scale. *Personality and Individual Differences*, *12*(3), 241–253. https://doi.org/10.1016/0191-8869(91)90110-W

Ahn, S., & Brown-Schmidt, S. (2020). Retrieval processes and audience design. *Journal of Memory and Language*, *115*, 104149. https://doi.org/10.1016/j.jml.2020.104149

Antony, J. W., Piloto, L., Wang, M., Pacheco, P., Norman, K. A., & Paller, K. A. (2018). Sleep Spindle Refractoriness Segregates Periods of Memory Reactivation. *Current Biology*, *28*(11), 1736-1743.e4. https://doi.org/10.1016/j.cub.2018.04.020

Ball, L. V., Kimel, E., Keller, V. G., Ward, E., Cairney, S. A., Mak, M. H. C., Li, L., Rodd, J. M., & Gaskell, M. G. (2025). No evidence for a targeted memory reactivation effect on word-meaning priming. *Neuropsychologia*, *219*, 109264. https://doi.org/10.1016/j.neuropsychologia.2025.109264

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using**lme4**. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Bayen, U. J., Murnane, K., & Erdfelder, E. (1996). Source discrimination, item detection, and multinomial models of source monitoring. *Journal of Experimental Psychology:*

*Learning, Memory, and Cognition*, *22*(1), 197–215. https://doi.org/10.1037/0278-7393.22.1.197

Berens, P. (2009). **CircStat**: A*MATLAB*Toolbox for Circular Statistics. *Journal of Statistical Software*, *31*(10). https://doi.org/10.18637/jss.v031.i10

Berres, S., & Erdfelder, E. (2021). The sleep benefit in episodic memory: An integrative review and a meta-analysis. *Psychological Bulletin*, *147*(12), 1309–1353. https://doi.org/10.1037/bul0000350

Berres, S., Erdfelder, E., & Kuhlmann, B. G. (2025). Does sleep benefit source memory? Investigating 12-h retention intervals with a multinomial modeling approach. *Memory & Cognition*. https://doi.org/10.3758/s13421-024-01579-8

Born, J., & Wilhelm, I. (2012). System consolidation of memory during sleep. *Psychol Res*, *76*(2), 192–203. https://doi.org/10.1007/s00426-011-0335-6

Brown, A. S., Jones, E. M., & Davis, T. L. (1995). Age differences in conversational source monitoring. *Psychology and Aging*, *10*(1), 111.

Brown-Schmidt, S. (2009). Partner-specific interpretation of maintained referential precedents during interactive dialog. *Journal of Memory and Language*, *61*(2), 171–190. https://doi.org/10.1016/j.jml.2009.04.003

Brown-Schmidt, S., & Duff, M. C. (2016). Memory and common ground processes in language use. In *Topics in Cognitive Science* (Vol. 8, Issue 4, pp. 722–736). Wiley Online Library.

Brown-Schmidt, S., Jaeger, C. B., Evans, M. J., & Benjamin, A. S. (2023). MEMCONS: How Contemporaneous Note-Taking Shapes Memory for Conversation. *Cognitive Science*, *47*(4), e13271. https://doi.org/https://doi.org/10.1111/cogs.13271

Brown-Schmidt, S., Yoon, S. O., & Ryskin, R. A. (2015). People as contexts in conversation. In *Psychology of learning and motivation* (Vol. 62, pp. 59–99). Elsevier.

Buysse, D. J., Reynolds 3rd, C. F., Monk, T. H., Berman, S. R., & Kupfer, D. J. (1989). The Pittsburgh Sleep Quality Index: A new instrument for psychiatric practice and research. *Psychiatry Res*, *28*(2), 193–213. https://doi.org/10.1016/0165-1781(89)90047-4

Clark, H. H., & Brennan, S. (1991). Grounding in communication. *Perspectives on Socially Shared Cognition*. https://api.semanticscholar.org/CorpusID:153811205

Clark, H. H., & Murphy, G. L. (1982). Audience Design in Meaning and Reference. In *Advances in Psychology* (Vol. 9, pp. 287–299). Elsevier. https://doi.org/10.1016/S0166-4115(09)60059-5

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, *22*(1), 1–39. https://doi.org/https://doi.org/10.1016/0010-0277(86)90010-7

Cordi, M. J., & Rasch, B. (2021). How robust are sleep-mediated memory benefits? *Current Opinion in Neurobiology*, *67*, 1–7. https://doi.org/10.1016/j.conb.2020.06.002

Cowan, E. T., Schapiro, A. C., Dunsmoor, J. E., & Murty, V. P. (2021). Memory consolidation as an adaptive process. *Psychonomic Bulletin & Review*, *28*(6), 1796–1810. https://doi.org/10.3758/s13423-021-01978-x

Denis, D., Bottary, R., Cunningham, T. J., Davidson, P., Yuksel, C., Milad, M. R., & Pace-Schott, E. F. (2025). Slow oscillation-sleep spindle coupling is associated with expectancy measures of fear extinction retention in trauma-exposed individuals. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, S2451902225002575. https://doi.org/10.1016/j.bpsc.2025.08.009

Denis, D., & Cairney, S. A. (2023). Neural reactivation during human sleep. *Emerging Topics in Life Sciences*, *7*(5), 487–498. https://doi.org/10.1042/ETLS20230109

Denis, D., Mylonas, D., Poskanzer, C., Bursal, V., Payne, J. D., & Stickgold, R. (2021). Sleep Spindles Preferentially Consolidate Weakly Encoded Memories. *The Journal of Neuroscience*, *41*(18), 4088–4099. https://doi.org/10.1523/JNEUROSCI.0818-20.2021

Dickerson, B. C., & Eichenbaum, H. (2010). The Episodic Memory System: Neurocircuitry and Disorders. *Neuropsychopharmacology*, *35*(1), 86–104. https://doi.org/10.1038/npp.2009.126

Diekelmann, S., & Born, J. (2010). The memory function of sleep. *Nature Reviews Neuroscience*, *11*(2), 114–126. https://doi.org/10.1038/nrn2762

Diekelmann, S., Paulus, F. M., & Krach, S. (2018). Sleep to be social: The critical role of sleep and memory for social interaction. *Behavioral and Brain Sciences*, *41*, e10. Cambridge Core. https://doi.org/10.1017/S0140525X17001327

Diekelmann, S., Wilhelm, I., & Born, J. (2009). The whats and whens of sleep-dependent memory consolidation. *Sleep Medicine Reviews*, *13*(5), 309–321. https://doi.org/10.1016/j.smrv.2008.08.002

Duff, M. C., & Brown-Schmidt, S. (2012). The hippocampus and the flexible use and processing of language. *Frontiers in Human Neuroscience*, *6*, 69.

Duff, M. C., Gupta, R., Hengst, J. A., Tranel, D., & Cohen, N. J. (2011). The Use of Definite References Signals Declarative Memory. *Psychological Science*, *22*(5), 666–673. https://doi.org/10.1177/0956797611404897

Duff, M. C., Hengst, J. A., Tranel, D., & Cohen, N. J. (2008). Collaborative discourse facilitates efficient communication and new learning in amnesia. *Brain and Language*, *106*(1), 41–54. https://doi.org/10.1016/j.bandl.2007.10.004

Ellenbogen, J. M., Hulbert, J. C., Stickgold, R., Dinges, D. F., & Thompson-Schill, S. L. (2006). Interfering with Theories of Sleep and Memory: Sleep, Declarative Memory, and Associative Interference. *Current Biology*, *16*(13), 1290–1294. https://doi.org/10.1016/j.cub.2006.05.024

Fernandez, L. M. J., & Lüthi, A. (2020). Sleep Spindles: Mechanisms and Functions. *Physiological Reviews*, *100*(2), 805–868. https://doi.org/10.1152/physrev.00042.2018

Gais, S., Mölle, M., Helms, K., & Born, J. (2002). Learning-Dependent Increases in Sleep Spindle Density. *The Journal of Neuroscience*, *22*(15), 6830–6834. https://doi.org/10.1523/JNEUROSCI.22-15-06830.2002

Gaskell, M. G. (2024). EPS mid-career prize: An integrated framework for the learning, recognition and interpretation of words. *Quarterly Journal of Experimental Psychology*, *77*(12), 2365–2384. https://doi.org/10.1177/17470218241284289

Gaskell, M. G., Cairney, S. A., & Rodd, J. M. (2019). Contextual priming of word meanings is stabilized over sleep. *Cognition*, *182*, 109–126. https://doi.org/10.1016/j.cognition.2018.09.007

Gordon, A. M., & Chen, S. (2014). The Role of Sleep in Interpersonal Conflict: Do Sleepless Nights Mean Worse Fights? *Social Psychological & Personality Science*, *5*(2), 168–175. https://doi.org/10.1177/1948550613488952

Gorman, K. S., Gegg-Harrison, W., Marsh, C. R., & Tanenhaus, M. K. (2013). What's learned together stays together: Speakers' choice of referring expression reflects shared experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(3), 843–853. https://doi.org/10.1037/a0029467

Halonen, R., Kuula, L., Makkonen, T., Kauramäki, J., & Pesonen, A.-K. (2021). Self-Conscious Affect Is Modulated by Rapid Eye Movement Sleep but Not by Targeted Memory Reactivation–A Pilot Study. *Frontiers in Psychology*, *12*, 730924. https://doi.org/10.3389/fpsyg.2021.730924

Heck, D. W., Arnold, N. R., & Arnold, D. (2018). TreeBUGS: An R package for hierarchical multinomial-processing-tree modeling. *Behavior Research Methods*, *50*(1), 264–284. https://doi.org/10.3758/s13428-017-0869-7

Helfrich, R. F., Mander, B. A., Jagust, W. J., Knight, R. T., & Walker, M. P. (2018). Old Brains Come Uncoupled in Sleep: Slow Wave-Spindle Synchrony, Brain Atrophy, and Forgetting. *Neuron*, *97*(1), 221-230.e4. https://doi.org/10.1016/j.neuron.2017.11.020

Hoddes, E., Dement, W., & Zarcone, V. (1972). The development and use of the stanford sleepiness scale. *Psychophysiology*, *9*(1), 150.

Horne, J. A., & Ostberg, O. (1976). A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. *International Journal of Chronobiology*, *4*(2), 97–110.

Horton, W. S. (2007). The influence of partner-specific memory associations on language production: Evidence from picture naming. *Language and Cognitive Processes*, *22*(7), 1114–1139. https://doi.org/10.1080/01690960701402933

Horton, W. S., & Gerrig, R. J. (2002). Speakers' experiences and audience design: Knowing when and knowing how to adjust utterances to addressees. *Journal of Memory and Language*, *47*(4), 589–606.

Horton, W. S., & Gerrig, R. J. (2005). The impact of memory demands on audience design during language production. *Cognition*, *96*(2), 127–142. https://doi.org/10.1016/j.cognition.2004.07.001

Hupet, M., Chantraine, Y., & Nef, F. (1993). References in conversation between young and old normal adults. *Psychology and Aging*, *8*(3), 339.

Iber, C., Ancoli-Israel, S., Chesson, A. L., & Quan, S. (2007). The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications. *Westchester, IL: American Academy of Sleep Medicine*.

Isaacs, E. A., & Clark, H. H. (1987). References in conversation between experts and novices. *Journal of Experimental Psychology: General*, *116*(1), 26.

Jacoby, L. L. (1978). On interpreting the effects of repetition: Solving a problem versus remembering a solution. *Journal of Verbal Learning and Verbal Behavior*, *17*(6), 649–667.

Jenkins, J. G., & Dallenbach, K. M. (1924). Obliviscence during Sleep and Waking. *The American Journal of Psychology*, *35*(4), 605. https://doi.org/10.2307/1414040

Johns, M. W. (1991). A new method for measuring daytime sleepiness: The Epworth sleepiness scale. *Sleep*, *14*(6), 540–545. https://doi.org/10.1093/sleep/14.6.540

Jurica, P. J., & Shimamura, A. P. (1999). Monitoring item and source information: Evidence for a negative generation effect in source memory. *Memory & Cognition*, *27*(4), 648–656.

Klinzing, J. G., Niethard, N., & Born, J. (2019). Mechanisms of systems memory consolidation during sleep. *Nature Neuroscience*, *22*(10), 1598–1610. https://doi.org/10.1038/s41593-019-0467-3

Knutsen, D., & Le Bigot, L. (2020). Estimating Each Other's Memory Biases in Dialogue. *Discourse Processes*, *58*(2), 155–176. https://doi.org/10.1080/0163853x.2020.1837541

Koriat, A., Ben-Zur, H., & Druch, A. (1991). The contextualization of input and output events in memory. *Psychological Research*, *53*, 260–270.

Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian Cognitive Modeling: A Practical Course* (1st edn). Cambridge University Press. https://doi.org/10.1017/CBO9781139087759

Lo, J. C., Dijk, D.-J., & Groeger, J. A. (2014). Comparing the Effects of Nocturnal Sleep and Daytime Napping on Declarative Memory Consolidation. *PLoS ONE*, *9*(9), e108100. https://doi.org/10.1371/journal.pone.0108100

MacLeod, C. M., Gopie, N., Hourihan, K. L., Neary, K. R., & Ozubko, J. D. (2010). The production effect: Delineation of a phenomenon. *Journal of Experimental*

*Psychology: Learning, Memory, and Cognition*, *36*(3), 671–685. https://doi.org/https://doi.org/10.1037/a0018785

Mak, M. H. C., Curtis, A. J., Rodd, J. M., & Gaskell, M. G. (2023). Episodic memory and sleep are involved in the maintenance of context-specific lexical information. *Journal of Experimental Psychology: General*, *152*(11), 3087–3115. https://doi.org/10.1037/xge0001435

Mak, M. H. C., Curtis, A. J., Rodd, J. M., & Gaskell, M. G. (2024). Recall and recognition of discourse memory across sleep and wake. *Journal of Memory and Language*, *138*, 104536. https://doi.org/10.1016/j.jml.2024.104536

Maranges, H. M., & McNulty, J. K. (2017). The rested relationship: Sleep benefits marital evaluations. *J Fam Psychol*, *31*(1), 117–122. https://doi.org/10.1037/fam0000225

McKinley, G. L., Brown-Schmidt, S., & Benjamin, A. S. (2017). Memory for conversation and the development of common ground. *Memory & Cognition*, *45*, 1281–1294.

Mikutta, C., Feige, B., Maier, J. G., Hertenstein, E., Holz, J., Riemann, D., & Nissen, C. (2019). Phase-amplitude coupling of sleep slow oscillatory and spindle activity correlates with overnight memory consolidation. *Journal of Sleep Research*, *28*(6), e12835. https://doi.org/10.1111/jsr.12835

Morey, R. D., & Rouder, J. N. (2026). *BayesFactor: Computation of Bayes Factors for Common Designs*. https://doi.org/10.32614/CRAN.package.BayesFactor

Muehlroth, B. E., Sander, M. C., Fandakova, Y., Grandy, T. H., Rasch, B., Shing, Y. L., & Werkle-Bergner, M. (2019). Precise Slow Oscillation–Spindle Coupling Promotes

Memory Consolidation in Younger and Older Adults. *Scientific Reports*, *9*(1), 1940. https://doi.org/10.1038/s41598-018-36557-z

Nault, D. R., Voleti, R., Nicastro, M., & Munhall, K. G. (2023). Investigating the influence of local and personal common ground on memory for conversation using an online referential communication task. *Journal of Experimental Psychology: General*, *152*(6), 1598.

Neske, G. T. (2016). The Slow Oscillation in Cortical and Thalamic Networks: Mechanisms and Functions. *Frontiers in Neural Circuits*, *9*. https://doi.org/10.3389/fncir.2015.00088

Ng, T., Noh, E., & Spencer, R. M. (2025). *Does slow oscillation-spindle coupling contribute to sleep-dependent memory consolidation? A Bayesian meta-analysis*. https://doi.org/10.7554/eLife.101992.2

Payne, J. D., Schacter, D. L., Propper, R. E., Huang, L.-W., Wamsley, E. J., Tucker, M. A., Walker, M. P., & Stickgold, R. (2009). The role of sleep in false memory formation. *Neurobiology of Learning and Memory*, *92*(3), 327–334. https://doi.org/10.1016/j.nlm.2009.03.007

Rasch, B., & Born, J. (2013). About Sleep's Role in Memory. *Physiological Reviews*, *93*(2), 681–766. https://doi.org/10.1152/physrev.00032.2012

Ray, L. B., Baena, D., & Fogel, S. M. (2024). "Counting sheep PSG": EEGLAB-compatible open-source matlab software for signal processing, visualization, event marking and staging of polysomnographic data. *Journal of Neuroscience Methods*, *407*, 110162. https://doi.org/10.1016/j.jneumeth.2024.110162

Riefer, D. M., Chien, Y., & Reimer, J. F. (2007). Positive and negative generation effects in source monitoring. *Quarterly Journal of Experimental Psychology*, *60*(10), 1389–1405.

Schabus, M., Gruber, G., Parapatics, S., Sauter, C., Klösch, G., Anderer, P., Klimesch, W., Saletu, B., & Zeitlhofer, J. (2004). Sleep Spindles and Their Significance for Declarative Memory Consolidation. *Sleep*, *27*(8), 1479–1485. https://doi.org/10.1093/sleep/27.7.1479

Schmidig, F. J., Ruch, S., & Henke, K. (2024). Episodic long-term memory formation during slow-wave sleep. *eLife*, *12*, RP89601. https://doi.org/10.7554/eLife.89601

Schnuerch, M., Erdfelder, E., & Heck, D. W. (2020). Sequential hypothesis tests for multinomial processing tree models. *Journal of Mathematical Psychology*, *95*, 102326. https://doi.org/10.1016/j.jmp.2020.102326

Schober, M. F., & Clark, H. H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology*, *21*(2), 211–232.

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory*, *4*(6), 592. https://doi.org/https://psycnet.apa.org/doi/10.1037/0278-7393.4.6.592

Sterpenich, V., van Schie, M. K. M., Catsiyannis, M., Ramyead, A., Perrig, S., Yang, H.-D., Van De Ville, D., & Schwartz, S. (2021). Reward biases spontaneous neural reactivation during sleep. *Nature Communications*, *12*(1), 4162. https://doi.org/10.1038/s41467-021-24357-5

Stickgold, R., & Walker, M. P. (2013). Sleep-dependent memory triage: Evolving generalization through selective processing. *Nature Neuroscience*, *16*(2), 139–145. https://doi.org/10.1038/nn.3303

Tamminen, J., Lambon Ralph, M. A., & Lewis, P. A. (2013). The Role of Sleep Spindles and Slow-Wave Activity in Integrating New Information in Semantic Memory. *The Journal of Neuroscience*, *33*(39), 15376–15381. https://doi.org/10.1523/JNEUROSCI.5093-12.2013

Tamminen, J., Payne, J. D., Stickgold, R., Wamsley, E. J., & Gaskell, M. G. (2010). Sleep Spindle Activity is Associated with the Integration of New Memories and Existing Knowledge. *The Journal of Neuroscience*, *30*(43), 14356–14360. https://doi.org/10.1523/JNEUROSCI.3028-10.2010

van der Helm, E., Gujar, N., & Walker, M. P. (2010). Sleep Deprivation Impairs the Accurate Recognition of Human Emotions. *Sleep*, *33*(3), 335–342. https://doi.org/10.1093/sleep/33.3.335

Van Schalkwijk, F. J., Sauter, C., Hoedlmoser, K., Heib, D. P. J., Klösch, G., Moser, D., Gruber, G., Anderer, P., Zeitlhofer, J., & Schabus, M. (2019). The effect of daytime napping and full-night sleep on the consolidation of declarative and procedural information. *Journal of Sleep Research*, *28*(1), e12649. https://doi.org/10.1111/jsr.12649

Voeten, C. C. (2023). *buildmer: Stepwise Elimination and Term Reordering for Mixed-Effects Regression*. https://CRAN.R-project.org/package=buildmer

Wilkes-Gibbs, D., & Clark, H. H. (1992). Coordinating beliefs in conversation. *Journal of Memory and Language*, *31*(2), 183–194. https://doi.org/https://doi.org/10.1016/0749-596X(92)90010-U

Yonelinas, A. P., Ranganath, C., Ekstrom, A. D., & Wiltgen, B. J. (2019). A contextual binding theory of episodic memory: Systems consolidation reconsidered. *Nature Reviews Neuroscience*, *20*(6), 364–375. https://doi.org/10.1038/s41583-019-0150-4

Yoon, S. O., & Brown-Schmidt, S. (2014). Adjusting conceptual pacts in three-party conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(4), 919.

Yoon, S. O., Duff, M. C., & Brown-Schmidt, S. (2017). Learning and using knowledge about what other people do and don't know despite amnesia. *Cortex*, *94*, 164–175. https://doi.org/10.1016/j.cortex.2017.06.020

Yoon, S. O., Duff, M. C., & Brown-Schmidt, S. (2024). Keeping track of who knows what in multiparty conversation despite severe memory impairment. *Neuropsychologia*, *194*, 108780. https://doi.org/10.1016/j.neuropsychologia.2023.108780

Zhang, J., Yang, Y., & Hong, Y.-Y. (2019). Sleep deprivation undermines the link between identity and intergroup bias. *Sleep*, zsz213. https://doi.org/10.1093/sleep/zsz213