

# **Emotive content and sleep enhance memory for metaphorical language**

**Chloe R. Newbury<sup>1</sup>, Padraic Monaghan<sup>1</sup>, Francesca M.M. Citron<sup>1\*</sup>**

<sup>1</sup>Department of Psychology, Lancaster University, UK

\*Corresponding author: Francesca M.M.Citron, Department of Psychology, Lancaster University, Lancaster LA1 4YF, UK. email: fmm.citron@gmail.com

Running head: Sleep alters emotion in metaphors

## Abstract

Memory for emotional information is greater than for non-emotional information, and is enhanced by sleep-related consolidation. Previous studies have focused on emotional arousal and valence of established stimuli, but what is the effect of sleep on newly acquired emotional information? Figurative expressions, which are pervasive in everyday communication, are often rated as higher in emotionality than their literal counterparts, but the effect of emotionality on the learning of metaphors, and the effect of sleep on newly acquired emotionally negative, positive, and neutral language is as yet poorly understood. In this study, participants were asked to memorise conventional (e.g., “*sunny disposition*”) and novel (e.g., “*cloudy disposition*”) metaphorical word pairs varying in valence, accompanied by their definitions. After a 12-hour period of sleep or wake, participants were tested on their recognition of word pairs and recall of definitions. We found higher arousal ratings were related to increased recognition and recall performance. Furthermore, sleep increased accurate recognition of all word pairs compared to wake, but also reduced the valence of word pairs. The results indicate better memory for newly acquired emotional stimuli, a benefit of sleep for memory, but also a reduction of emotional arousal as a consequence of sleep consolidation.

Key words: sleep, memory, emotion, language learning, novel metaphor

## 1. Introduction

Figurative language is defined as a form of language use where the meaning of the stimulus is not directly available from the words it comprises, but is learned by convention. For example “bright student” describes the intelligence, rather than the luminosity, of the person.

Figurative expressions are pervasive in everyday communication (Cameron, 2008), yet they are unique as linguistic expressions in that they tend to be emotionally charged, far beyond the emotionality of the words they contain (Citron & Goldberg, 2014; Citron et al., 2020a). In this study, we uncover some relevant processes involved in the acquisition of figurative language, and unpack how their emotional content unfolds.

Conceptual Metaphor Theory (Lakoff & Johnson, 1980) suggests that metaphors typically consist of mappings between concrete and abstract concepts. For example ‘*bright*’, something we can experience with our senses, is used metaphorically to convey the otherwise abstract concept ‘*intelligent*’. Metaphorical language is used much more frequently to describe personal events that are high in emotional intensity (Fainsilber & Ortony, 1987). Furthermore, Citron and Goldberg (2014) found that conventional metaphors conveying taste were more emotionally engaging than literal sentences and increased activation of the left hippocampus and amygdala; similar results were found for longer passages containing metaphors not restricted to taste (Citron et al., 2016; see also Citron et al., 2020b). The amygdala has been associated with emotional memory processing (McGaugh, 2004), and the encoding of emotional memories has been correlated with co-activation of both the amygdala and hippocampus (for a review, see Walker, 2021).

Increased activation in a number of brain regions (see meta-analysis by Bohrn et al., 2012) suggests that greater processing resources are involved in the understanding of figurative compared to literal language. However, conventional metaphors (e.g., “*bright*

*student*”) have been found to require much less processing resources than novel metaphors (Lai & Curran, 2013); in fact, conventional metaphors are processed as quickly as literal expressions as their meaning is accessed directly (Gibbs et al., 2004). The meaning of novel metaphors, by contrast, needs to be created on the fly; for example “her husband is an elephant” can mean that he is clumsy or large or that he has a long nose, and this requires inferences, theory of mind, and other pragmatic processes (Ferstl, 2010). Novel metaphorical word pairs (e.g., “*chilly awkwardness*”) may also differ in their sensibility, with some definitions (“uncomfortable social situation”) matching the novel word pair more congruently than others (“busy traffic with angry drivers”), and these congruent pairs might, again, be easier to process.

Though we know little about when and how figurative language engages readers or listeners emotionally, there is substantial knowledge of how emotion and memory inter-relate. For instance, verbal memory for emotional information tends to be more accurate than that of non-emotional information (Adelman & Estes, 2013; Kensinger & Corkin, 2003; McGaugh, 2004). Emotional valence, which describes how positive or negative a stimulus is, and arousal, relating to how calm or exciting an experience is (Russell, 2003), have different effects on memory. A meta-analysis by Pereira et al. (2023) on (non-verbal) emotional memories found that positively or negatively valenced memories were remembered less well than neutral memories. However, arousal (regardless of valence) resulted in improved memory compared to neutral memories. For verbal stimuli, the few studies that have distinguished valence and arousal show conflicting effects. Kensinger and Corkin (2003) found that words high in negative valence and high in arousal lead to the greatest increases in memory performance, whereas Adelman and Estes (2013) reported significant increases in memory performance for words of both negative and positive valence, however no effect of arousal. Thus, although differences in memory performance for stimuli varying in

emotionality are observed, the exact effects of valence and arousal for encoding verbal stimuli remain somewhat unclear.

A range of evidence indicates a role of sleep in promoting both learning and memory (Berres & Erdfelder, 2021; Cellini et al., 2016; Cordi & Rasch, 2021; Rasch & Born, 2013), and sleep may thus be vital to the encoding of new figurative expressions, and their emotional enkindling. For example, a period of sleep after learning increases memory for word pairs compared to wakefulness (Wilson et al., 2012), and integration of novel word pairs into long-term memory has also been found to occur after a period of sleep (Dumay & Gaskell, 2007). The Active Systems Consolidation hypothesis (Diekelmann & Born, 2010; Marshall & Born, 2007) suggests that this boost to learning and memory during offline periods of sleep is due to the reactivation and integration of these memories from the hippocampus, where they are initially encoded, into long-term semantic representations in the neocortex. This encoding in the cortex can then reduce effects of interference that may occur if encoding remains in the hippocampus (Dastgheib et al., 2022; Frankland & Bontempi, 2005).

In addition to promoting memory of information content, sleep may affect the processing of emotional information (Chambers & Payne, 2014; Niu et al., 2024; Payne et al., 2008; Reid et al., 2022; Wagner et al., 2001; Walker, 2021). Lipinska et al. (2019) conducted a meta-analysis of sleep effects on emotional memory consolidation, and found that there was no overall effect of sleep in promoting memory for emotional stimuli in particular, though there was some evidence of a beneficial effect when studies involved overnight sleep or free recall instead of only recognition memory. Focusing on emotional memory recognition studies, Schäfer et al. (2020) found in a meta-analysis that sleep improved memory for neutral and emotional stimuli compared to wake. There was no overall benefit for emotional stimuli from sleep, but there was some evidence that REM sleep might relate to a boost to

emotional memory. Davidson et al. (2023) conducted a meta-analysis of studies comparing sleep and wake groups on emotional memory consolidation. They found that the majority of studies did not find an effect of sleep on emotional memory compared to neutral stimuli, and concluded that, if it is there, then it is likely a small effect.

In parallel with consolidating memory content, sleep may also change the affective tone of the information itself. The sleep to forget, sleep to remember (SFSR) hypothesis (Walker, 2021) suggests that emotional memories do persist over time, but the emotional response to such information is manipulable (e.g., Xia et al., 2023) and can be *reduced* as a consequence of sleep, though such effects are not always observed (e.g., Cellini et al., 2016). However, a key feature of previous stimuli used in sleep and memory studies of emotion is that they employ stimuli that are already imbued with emotion, which prevents an investigation of how emotion and memory might develop in synchrony. In this study, therefore, we provided participants with novel stimuli – metaphorical expressions – that become charged with emotion during learning.

We also varied the extent to which the meaning of the figurative expressions to be learned were congruent or incongruent with the words they contained. Figurative language varies in this respect, for instance “*She spilled the beans*” is congruent (or *semantically transparent*), with its idiomatic meaning of ‘revealing a secret’ which can be almost intuitively inferred from its constituting words; whereas “*He went down rabbit holes*” is incongruent (or *semantically opaque*) in that the words it contains are unrelated to its idiomatic meaning of ‘becoming deeply involved in a complex situation, which often leads to even more complications’; in fact, this meaning must be learned (Keysar & Bly, 1995). Typically, semantically transparent figurative expressions are processed more quickly, and thus easily, than opaque expressions (e.g., Bell & Schäfer, 2016; Gibbs, Nayak & Cutting, 1995). This

congruence, or semantic transparency, might also relate to different effects of sleep on learning and memory. Payne et al. (2012) compared learning of word pairs that were either semantically related or unrelated, and found that sleep resulted in a larger advantage for memory of unrelated word pairs. Ashton et al. (2022) found that semantically incongruent word pairings (e.g., red-elephant) were remembered worse than congruent pairings, but this disadvantage was significantly smaller following sleep, indicating that sleep promoted storage of the incongruent pairs. Sio et al. (2013) found, for a problem solving study, that harder-to-process material benefited more from sleep than wake, consistent with the observation that unrelated material – also harder to learn – is boosted by sleep.

The current study therefore investigates whether sleep differentially affects memory for conventional and novel metaphorical word pairs of negative, positive, and neutral valence that vary in their congruence. We had four predictions from our study. First, there will be an overall benefit of sleep compared to wake on both recognition and definition recall accuracy of both conventional and novel metaphors as sleep promotes memory. Second, sleep will enhance memory for the harder-to-process novel metaphorical word pairs compared to conventional metaphors, and also that sleep will boost incongruent more than congruent pairs. Third, emotionality of the word pairs will enhance their accurate recognition (Kensinger & Corkin, 2003; Adelman & Estes, 2013). Due to inconsistencies within the literature as to the effects of valence versus arousal, we may expect to see an increase in memory performance for word pairs rated high in negative valence and arousal (Kensinger & Corkin, 2003), or increases in memory for word pairs of both negative and positive valence, independent of arousal (Adelman & Estes, 2013). Fourth, if sleep affects emotional memory processing (Walker, 2021), then, according to the SFSR hypothesis, ratings of arousal and valence of the word pairs would be reduced as a consequence of sleep.

## 2. Method

### 2.1 Participants

Forty-nine Lancaster University students took part in the study for course credit or payment of £7. All participants gave informed consent and were fully debriefed at the end of the study. Non-native British English speakers ( $n = 6$ ) were removed from the analysis, as well as one participant who self-reported being in bed for less than 6 hours (cut-off based on previous studies of sleep and memory, e.g., Diekelmann et al., 2008). Two participants were removed due to computer error during completion of the experiment. This led to the inclusion of 40 participants (34 female), *mean age* = 20.8 years, *SD* = 3.01, *range* = 18-32, which meant we had power of 0.8 to find effects of  $d = 0.8$  or greater. Participants were randomly allocated to the sleep,  $N = 20$ , 5 female, 15 male, *mean age* = 20.4 years, *SD* = 1.6, *range* = 18-24, or wake condition,  $N = 20$ , 1 female, 19 male, *mean age* = 21.2 years, *SD* = 3.9, *range* = 18-32. Ethical approval was granted by the University's faculty ethics board, and the study was conducted in accordance with World Medical Association Declaration of Helsinki ethical guidelines.

### 2.2 Materials

We collated 71 conventional and 67 novel metaphorical word pairs from a range of sources: 56 conventional and 59 novel word pairs were taken from Liu, Connell and Lynott (2019), two conventional word pairs from Mashal and Itkes (2012), five conventional and four novel word pairs from Forgács, Lukás, and Pléh (2014). Eight conventional and four novel metaphors were also added.



Using these word pairs, we collected ratings on a 7-point scale measuring sensibility (the extent to which a word pair made sense), metaphoricity (the extent to which a word pair expressed a non-literal expression), familiarity (extent of exposure to the word pair), emotional valence (from negative to positive), and emotional arousal (from non-arousing to arousing), from 15 participants who did not take part in the main experiment. Thirteen of those same participants completed a second questionnaire providing a definition of what they thought the meaning of each novel word pair was.

We extracted 30 conventional metaphorical word pairs by choosing those that were rated high in metaphoricity (mean = 4.97, SD = 0.68), sensibility (mean = 6.39, SD = 0.33), and familiarity (mean = 5.48, SD = 0.76). We also extracted 60 novel metaphorical word pairs rated low in familiarity (mean = familiarity (mean = 2.15, SD = 0.83, comparison to conventional metaphors:  $t(62.94) = 18.98, p < .001, d = 4.12$ ), to ensure these were not previously known to participants. The novel word pairs were also rated lower in metaphoricity (mean = 4.61, SD = 0.57, comparison to conventional metaphors:  $t(50.30) = 2.50, p = .016, d = 0.59$ ), and sensibility (mean = 3.88, SD = 1.09, comparison to conventional metaphors:  $t(77.55) = 16.35, p < .001, d = 2.75$ ). For both conventional and novel word pairs, we ensured that a third were rated high in negative valence (mean = -1.74, SD = 0.49), a third high in positive valence (mean = 1.44, SD = 0.54), and a third neutral in valence (mean = -0.14, SD = 0.49), which were significantly different than the negative,  $t(58.00) = -12.71, p < .001, d = -3.28$ , and the positive,  $t(57.34) = 11.76, p < .001, d = 3.04$ , word pairs. We chose negative and positive word pairs with ratings of arousal that were as similar as possible (mean for negative = 3.57, SD = 0.84, mean for positive = 4.13, SD = 1.03,  $t(55.69) = 2.28, p = .026, d = 0.59$ ), and very different than the neutral word pairs which were selected to be low in arousal (mean = 2.43, SD = 0.98), comparison to negative word

pairs:  $t(57.85) = 6.53, p < .001, d = 1.69$ , comparison to positive word pairs:  $t(57.34) = 11.76, p < .001, d = 3.04$ .

We then divided these pairs into two sets, with 15 conventional and 30 novel word pairs each. One set was presented to participants during training and testing (seen pairs), the other set was reserved as an unseen set for testing (unseen pairs).

Within each set, the novel word pairs were divided into two categories: 15 word pairs in which we assigned a meaning that was congruent, i.e., similar to the meanings provided by participants in the questionnaire, and therefore sensible: e.g., “*jingling satisfaction*” was paired with the meaning “feeling very pleased and proud of yourself”; and 15 word pairs that were assigned an incongruent meaning, i.e., a meaning that differed from those provided by participants, and therefore less sensical: e.g., “*jingling satisfaction*” was paired with “a generous reward”. Whether a word pair was in the congruent or incongruent condition was counterbalanced: half of participants saw one set of word pairs with a congruent meaning, the other half saw the same set with an incongruent meaning. Conventional metaphors, instead, were always assigned their congruent, conventionalised meaning, e.g., “*handsome profit*” was paired with “a large sum of money made”.

The three metaphor conditions (conventional, novel congruent, novel incongruent) were each composed of three groups differing in valence, so that five word pairs had negative valence, five word pairs had positive valence, and five were neutral.

For seen word pairs, we ensured that conventional and novel word pairs differed in familiarity and sensibility, but were controlled as far as possible on valence and arousal. We also ensured that conventional and novel word pairs did not differ in metaphoricity, valence, or arousal, with the exception of positive conventional pairs which were significantly higher in metaphoricity and valence than positive novel pairs. Since metaphors are often highly

emotive (Fainsilber & Ortony, 1987; Citron et al., 2020a), this could be due to lower metaphoricity leading to lower valence ratings. We also checked that valence was significantly different for positive versus neutral versus negative lists for both conventional and novel word pairs. All statistical tests are reported in the Supplementary Information (SI, available at <https://osf.io/s3g98>).

For seen versus unseen word lists, we ensured that there were no overall differences in familiarity, sensibility, metaphoricity, valence, and arousal ratings for each subset of positive, neutral, and negative conventional word pairs. We also tested for any differences for the seen versus unseen novel word pairs. There were no significant differences except that negative seen novel word pairs were rated lower than unseen for metaphoricity, and positive seen novel word pairs were rated higher than unseen for valence (see Supplementary Information S1 for statistics). These differences are not problematic because effects of word pair properties on memory were only investigated within seen word pairs.

## **2.3 Procedure**

Participants in the wake group were exposed to the word pairs and definitions at 9 am and then tested the same day at 9 pm after a day of wakefulness. Participants in the sleep group were exposed to word pairs and definitions at 9 pm, and then tested at 9 am after a night of sleep. Participants in the sleep condition were required to wear an Actigraph sleep monitor overnight to measure time spent asleep.

During the first session, participants first completed a questionnaire on their general sleep habits, on their mood using the Self-Assessment Manikin (SAM; Bradley & Lang, 1994) questionnaire, which includes measures of pleasure (or emotional valence), activation

(or emotional arousal), and control (or dominance), and the Stanford Sleepiness Scale (SSS; Hoddes, Zarcone, Smythe, Phillips, & Dement, 1973). Participants were then seated approximately 60cm from the computer screen and were exposed to the 15 conventional, 15 novel congruent, and 15 novel incongruent word pairs and definitions. Participants were presented with each word pair and definition one at a time for 5 s each. The definition would then disappear from the screen, and participants were required to provide a rating of valence and arousal for the word pair. Valence was rated on a scale from negative (1) to positive (7), with 4 meaning completely neutral, and arousal was rated on a scale from not at all arousing (1) to completely arousing (7). Participants were asked to press the corresponding key on the keyboard to give their response. Participants were then required to reproduce the definition of the word pair by typing the meaning into the computer using the keyboard (First recall of definitions). Participants were not given a time limit to complete the ratings or definitions. Each word pair and definition were presented in random order and once only.

After all 45 word pairs and definitions had been presented to participants, we again gave participants each word pair one at a time in random order, without accompanying definition, and asked them to type the definition of each word pair again (Second recall). Participants were instructed not to guess, and to type “I don’t know” if they could not remember the definition. Each word pair remained on the screen until participants entered their definition, and participants were not given a time limit to complete these definitions. This test was included in order to determine the extent to which encoding of the definitions had been effective.

Participants then left the lab, and those in the wake condition were instructed not to nap throughout the day. Twelve hours later, after a period of overnight sleep or daytime wake, participants returned for the testing phase of the experiment. In this second session,

participants were again asked to complete the sleep habits questionnaire, the SAM and the SSS scales. Participants were then asked to complete a recognition task on the screen (same setup as for the recall task above). They were presented with a central fixation cross for 500 ms, before one of the word pairs was presented at the centre of the screen. Participants were asked to decide whether they recognised the word pair as one that was presented to them in the first session (*old*) or not presented in the first session (*new*). Participants made their decision on whether the word pair was old or new on a scale from one to six based on their confidence, with 1 indicating “definitely old”, and 6 indicating “definitely new” (see Weidemann & Kahana, 2016, for a similar approach). The word pair remained on the screen until participants made a decision. Participants were not given a response deadline but were asked to respond as quickly and accurately as possible. Once participants had made their decision, they were then presented with another fixation point for 500 ms, before another word pair appeared on the screen. The presentation order of each word pair was randomised for each participant.

After participants had seen all word pairs, they were then given each of the previously seen word pairs again, one at a time, without accompanying definition. Participants were first asked to rate the valence and arousal of the word pair using the same rating scales as in the first session, and then to provide the definition of the word pair by typing into the box provided on the computer screen (Third recall). Participants were instructed not to guess, and to type “I don’t know” if they did not know the answer. The word pair remained on the screen until participants gave the ratings and definition, and participants were not given a response deadline.

The procedure is summarised in Figure 1.

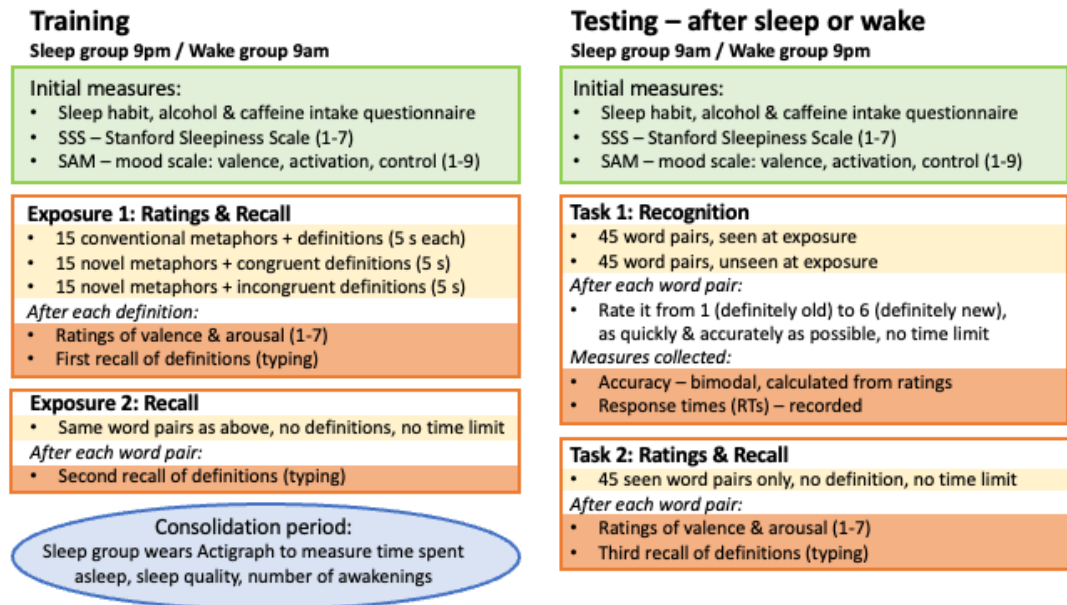


Figure 1: Experimental procedure.

## 2.4 Data Analysis

Data were analysed using the package lme4 in R. For binary outcomes (recall and recognition accuracy), generalised linear mixed-effects models were used. For response times (RTs), linear mixed-effects analyses were performed. For all analyses, to determine the best-fitting model, fixed effects were added incrementally. The model was initially fitted with random effects of participants and items (word pairs), and then fixed effects were added to the model one at a time. Each model was compared to the previous model using likelihood ratio tests, and fixed effects remained in the final model if they contributed significantly to model fit.

For recall of definitions, whether the definition provided by participants was correct or incorrect was determined by two scorers independently. Responses were scored as correct if the definition was the same or similar to that learned during training. If synonyms of the

definition were used, the definition was scored as correct. Where the scorers differed in their decision, the participant's response was discussed between scorers, and a final decision was made. Both recall and recognition accuracy were coded such that a correct response was given a score of 1, and an incorrect response was given a score of 0.

We converted the six-point confidence ratings (from definitely not seen, to definitely seen) to accuracies by recording accurate responses as those where the word pair was previously seen when participants judged it to have been seen before (with any degree of confidence, i.e., 1, 2, or 3 on the confidence scale) and when the word pair was not previously seen judged as not seen before, again with any degree of confidence (4, 5, or 6 on the confidence scale). Supplementary Information S2 reports analyses of these confidence ratings, which show very similar results to the conversion to accuracy responses reported in this paper.

RTs for correct responses in the recognition task were analysed. RTs exceeding 2.5 standard deviations from the mean were removed from the analysis, leading to the removal of 2.44% of the data, and leaving 3157 data points. Based on test of normality, we reduced skewness in the RT distribution by transforming latencies to  $\log_{10}(\text{RT})$ , and analysed the  $\log\text{RTs}$ , based on suggestions from Baayen, Feldman, and Schreuder (2006).

To compare performance between novel and conventional word pairs, and between novel congruent and novel incongruent word pairs, we coded the "type of metaphor" categorical variable using Helmert contrasts. For novel versus conventional metaphors (contrast A), conventional word pairs = 1, and both novel congruent and novel incongruent word pairs = -.5. To compare novel congruent and incongruent word pairs (contrast B), conventional word pairs = 0, novel congruent word pairs = -.5, and novel incongruent word pairs = .5. For the other categorical fixed effects: for the sleep versus wake group, sleep was

the reference category, for time of testing, the testing session was the reference category against which the training session was compared.

### **3. Results**

Data and analyses are available at: <https://osf.io/s3g98>. In order to determine the effect of sleep and wake consolidation on metaphor encoding and processing, we first ensured that sleep versus wake effects were not due to time-of-day effects (3.1). We then assessed recall accuracy of word pairs at the testing session (3.2), then we assessed change from training to testing sessions for recall accuracy (3.3), and for participants' ratings of emotional valence and emotional arousal (3.4), in order to determine if sleep affected changes in recall and emotionality over time. We focused on the effect of stimulus type on processing for the previously seen stimuli, as type of stimulus was only controlled for these items. The Supplementary Information contains reports of model construction and comparisons, as well as additional analyses of seen versus unseen stimuli (S1), and analyses of response times (S3).

#### **3.1 Analyses of time of day effects on performance**

##### *Participants' mood and sleepiness scores in each session*

Mean scores of sleepiness, mood valence, mood activation/arousal, and mood control/dominance (the latter three measures from the SAM) in training and testing sessions are reported in Table 1. We tested whether sleepiness or mood might be different for the sleep versus wake groups, due to time of day of the training session, also reported in Table 1. We found no significant effects apart from wake group mood activation scores being higher than



for the sleep group before the testing session. To determine whether this difference in mood activation had an influence on definition recall, we conducted a generalised linear mixed effect model with recall accuracy of definitions of word pairs at testing as the dependent variable. Adding mood activation scores did not improve model fit compared to a model with no fixed effects ( $\chi^2(1) = 1.125, p = .289$ ). We also tested whether mood activation scores affected word pair recognition at testing, and again adding it to a model with no fixed effects did not improve model fit ( $\chi^2(1) = 0.467, p = .495$ ). Thus, there was no evidence that differences in sleep or mood between the two groups at different times of day may have affected the results.

Table 1. Mean, SD and t-tests comparing wake and sleep groups for sleepiness, and SAM mood self-ratings.

Measure	Training Session				Testing Session			
	Sleep	Wake			Sleep	Wake		
	Mean (SD)	Mean (SD)	<i>t</i>	<i>p</i>	Mean (SD)	Mean (SD)	<i>t</i>	<i>p</i>
Sleepiness	3.56 (2.77)	2.77 (1.44)	1.88	.068	2.94 (1.08)	3.27 (1.36)	-0.83	.411
Mood	6.11 (1.20)	6.41 (1.23)	-0.75	.456	6.17 (1.38)	6.41 (1.85)	-0.46	.647
valence								
Mood	3.56 (1.30)	3.55 (1.20)	0.02	.981	3.50 (1.30)	4.55 (1.67)	-2.17	.037
Activation/								
Arousal								

Mood	4.94 (1.58)	5.59 (1.34)	-1.34	.189	5.39 (1.64)	5.55 (1.53)	-0.13	.900
Control/Do minance								

### *Time of Day Controls in Training Session*

**Emotional valence and arousal ratings of word pairs at training.** We analysed differences in participants' ratings of valence and arousal in the training session, to determine whether time of day affected ratings of the stimuli. Table 2 reports the means and SDs of ratings by sleep and wake group. Adding the fixed effect of group to a linear mixed effects model of valence ratings of the word pairs did not significantly improve model fit, indicating that there was no evident effect of time of day on the ratings,  $\chi^2(1) = 0.036$ ,  $p = .849$ ). Adding the fixed effect of group to a linear mixed effects model of arousal ratings of the word pairs also did not significantly improve model fit,  $\chi^2(1) = 0.803$ ,  $p = .370$ ).

**First recall of definitions of word pairs at training.** To check whether the two groups were different in the first session, indicating an effect of time of day on encoding, we ran a generalised linear mixed-effects model, with first recall of definition accuracy at the training session as the dependent variable (see Table 2 for descriptives). Adding a fixed effect of group to a model with only random effects did not improve model fit ( $\chi^2(1) = 1.202$ ,  $p = .273$ ), indicating, as expected, no significant difference between the sleep and wake groups.

**Second recall of definitions at training.** During the training session we also measured participants' ability to recall definitions for a second time, to gain a measure of recall accuracy at the end of the training session after a short retention period. Although all participants provided definitions for a second time during training, due to computer error only half of the data was collected (10 participants in the wake group, 10 in the sleep group). We thus ran a

generalised linear mixed-effects model on data from these 20 participants, with random effects of participants and items (see Table 2 for descriptives). Adding the effect of group to a model with only random effects did not significantly improve model fit, indicating no time-of-day effect on recall of definitions, ( $\chi^2(1) = 0.626, p = .429$ ).

Table 2. Means and standard deviations of ratings of valence and arousal; and first and second recall of definitions at training (scores indicate proportion of correct items recalled).

	Sleep	Wake
	Mean (SD)	Mean (SD)
Emotional valence	3.98 (2.05)	3.97 (2.02)
Emotional arousal	4.33 (1.63)	4.44 (1.65)
First recall of definitions	0.97 (0.17)	0.95 (0.21)
Second recall of definitions	0.54 (0.50)	0.58 (0.49)

### 3.2 Effects of sleep on recognition of word pairs at testing

As fixed effects, we coded stimulus type as Helmert contrasts (comparing conventional versus novel word pairs, and whether the novel pairs were congruent or incongruent). We then included the comparison between positive and neutral stimuli and between negative and neutral stimuli, as well as actual emotional valence and arousal ratings of word pairs given by participants during the training session to determine whether individual ratings affected processing. We then added the fixed effect of group (sleep or

wake) as a main effect, and in interaction with stimulus type and emotionality. Effects were only retained in the model if they contributed significantly to model fit.

For recognition accuracy, the final model included a significant fixed effect of congruent vs. incongruent novel metaphors ( $\chi^2(1) = 5.575, p = .018$ ): participants recognised novel congruent word pairs (mean = 0.89, SD = 0.13) better than novel incongruent metaphors (mean = 0.84, SD = 0.12), contrary to expectations. The model also included arousal ratings at training ( $\chi^2(1) = 5.362, p = .021$ ), with greater arousal showing higher accuracy; and wake vs. sleep group ( $\chi^2(1) = 7.058, p = .008$ ): as predicted, the sleep group (mean = 0.91, SD = 0.05) outperformed the wake group (mean = 0.84, SD = 0.10). No other main effects or interactions significantly contributed to model fit, and the final model is reported in Table 3.

Table 3. Summary of fixed effects in final model of recognition accuracy of previously seen word pairs.

Recognition Accuracy					
<i>Predictors</i>	<i>Odds Ratios std. Error CI</i>			<i>Z</i>	<i>p</i>
(Intercept)	8.36	2.45	4.70 – 14.86	7.23	< .001
Congruent vs. incongruent word pair	1.48	0.26	1.04 – 2.10	2.20	.028
Arousal at training	1.12	0.05	1.02 – 1.24	2.41	.016

Group [wake vs. sleep]	0.49	0.12	0.30 – 0.80	-2.83	.005
Observations	1800, 40 participants, 45 word pairs				
Marginal R <sup>2</sup> / Conditional R <sup>2</sup>	0.045 / 0.229				

### 3.3 Effects of sleep on recall of definitions at training (Second recall) vs. testing

We next modelled participants' ability to recall definitions of word pairs at training vs. testing. The final model (see Table 4) included a significant fixed effect of time of testing ( $\chi^2(1) = 7.789, p = .005$ ): with reduced accuracy at testing (mean proportion correct = 0.45, SD = 0.14) compared to training (mean = 0.56, SD = 0.11), as anticipated due to a longer period between exposure and testing; an effect of conventional vs. novel metaphors ( $\chi^2(1) = 33.782, p < .001$ ): conventional (mean = 0.73, SD = 0.23) recalled better than novel (mean = 0.33, SD = 0.13); congruent vs. incongruent novel metaphors ( $\chi^2(1) = 641.620, p < .001$ ): congruent novel (mean = 0.59, SD = 0.23) better recalled than incongruent novel (mean = 0.07, SD = 0.08); and valence ratings at training ( $\chi^2(1) = 6.242, p = .012$ ): word pairs rated as more negative were more accurately recalled. No other main effects or interactions significantly contributed to model fit.

Table 4. Summary of fixed effects in final model of recall of definitions, comparing training and testing sessions.

<i>Predictors</i>	<i>Odds</i>	<i>std.</i>	<i>CI</i>	<i>z</i>	<i>p</i>
	<i>Ratios</i>	<i>Error</i>			

(Intercept)	0.88	0.24	0.51 – 1.50	-0.48	0.631
Training vs. testing	1.54	0.21	1.18 – 2.01	3.17	.002
Conventional v. novel word pairs	6.23	1.27	4.18 – 9.29	8.98	<.001
Congruent v. incongruent novel word pairs	41.05	7.55	28.62 – 58.88	20.18	<.001
Valence rating at training	0.90	0.04	0.82 – 0.98	-2.48	.013
Observations	2700, 40 participants, 45 word pairs				
Marginal R <sup>2</sup> / Conditional R <sup>2</sup>	0.446 / 0.633				

### 3.4 Effects of change in emotion for stimuli

When valence ratings, including ratings both at training and testing, were the outcome variable, emotionality of the word pairs significantly contributed to model fit ( $\chi^2(2) = 87.049$ ,  $p < .001$ ), confirming that positive word pairs were rated as more positive and negative word pairs more negative than neutral word pairs; in addition, arousal ratings at training significantly contributed to model fit ( $\chi^2(1) = 91.871$ ,  $p < .001$ ): word pairs rated as more arousing at training were rated as more positive. The interaction between time of testing and conventional vs. novel metaphor was also significant ( $\chi^2(2) = 8.238$ ,  $p = .016$ ): while conventional metaphors declined in positive ratings from training to testing, novel metaphors increased in positive ratings (see Figure 2). The interaction between time of testing and

emotionality of the word pair was significant ( $\chi^2(2) = 33.374, p < .001$ ): at testing, positive word pairs were rated as less positive, negative word pairs were rated as less negative, and neutral pairs remained the same than at training (see Figure 3). Finally, the interaction between time of testing and arousal (at training) was significant ( $\chi^2(1) = 8.413, p = .004$ ): increasingly more arousing pairs were rated as less positive at testing than training. Critically, there was also a significant interaction between time of testing and sleep or wake group ( $\chi^2(2) = 7.443, p = .024$ ): sleep reduced valence ratings of word pairs ( $z = 3.46, p = .003$ ), whereas there was no evidence of an effect for wake ( $z = -0.23, p = .996$ ), see Figure 4, and final model in Table 5.

Table 5. Summary of fixed effects in final model of valence ratings across training and testing sessions.

<i>Predictors</i>	<i>Estimates</i>	<i>std. Error</i>	<i>CI</i>	<i>t</i>	<i>p</i>
(Intercept)	3.39	0.16	3.08 – 3.70	21.53	< .001
Emotionality [negative vs. neutral]	-1.82	0.21	-2.23 – -1.42	-8.82	< .001
Emotionality [positive vs. neutral]	1.74	0.21	1.33 – 2.15	8.29	< .001
Arousal rating at training	0.17	0.02	0.13 – 0.20	10.35	< .001
Novel vs. conventional word pairs x Training	0.32	0.12	0.09-0.56	2.70	.007
Novel vs. conventional word pairs x Testing	0.24	0.12	0.01 – 0.48	2.03	.042

Emotionality Training	[negative]	×	0.38	0.09	0.21 – 0.55	4.40	< .001
Emotionality Testing	[positive]	×	-0.06	0.10	-0.26 – 0.13	-0.64	.525
Arousal rating at training Training vs. Testing session		×	-0.06	0.02	-0.09 – -0.03	-3.82	< .001
Sleep vs. wake group at Training			-0.13	0.07	-0.26 – 0.01	-1.79	.074
Sleep vs. wake group at Testing			0.07	0.07	-0.07 – 0.21	1.00	.318
Observations			3600, 40 participants, 90 word pairs				
Marginal R <sup>2</sup> / Conditional R <sup>2</sup>			0.578 / 0.660				



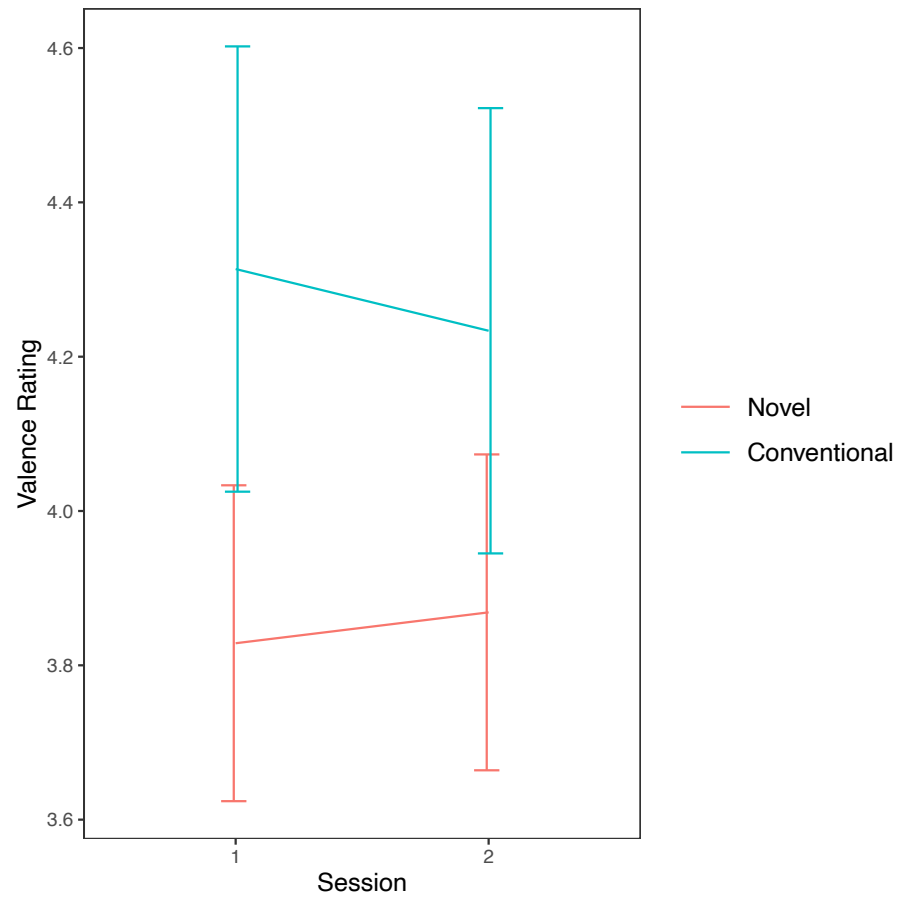


Figure 2. Mean valence ratings and 95% confidence intervals at training (1) vs. testing (2) for conventional and novel metaphorical word pairs.

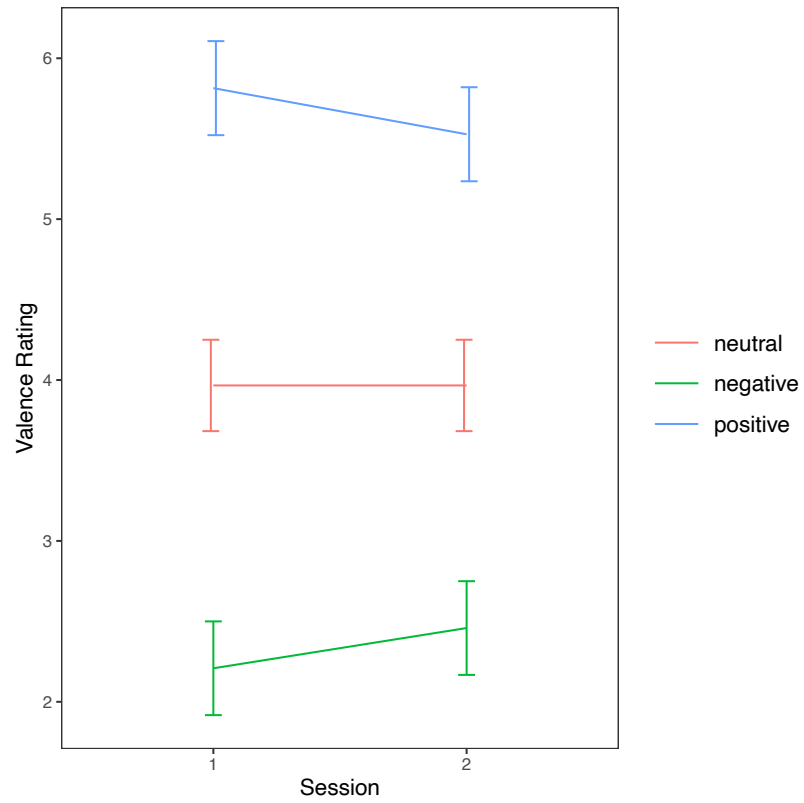


Figure 3. Mean valence ratings and 95% confidence intervals at training (1) vs. testing (2) for word pairs with different emotionality.

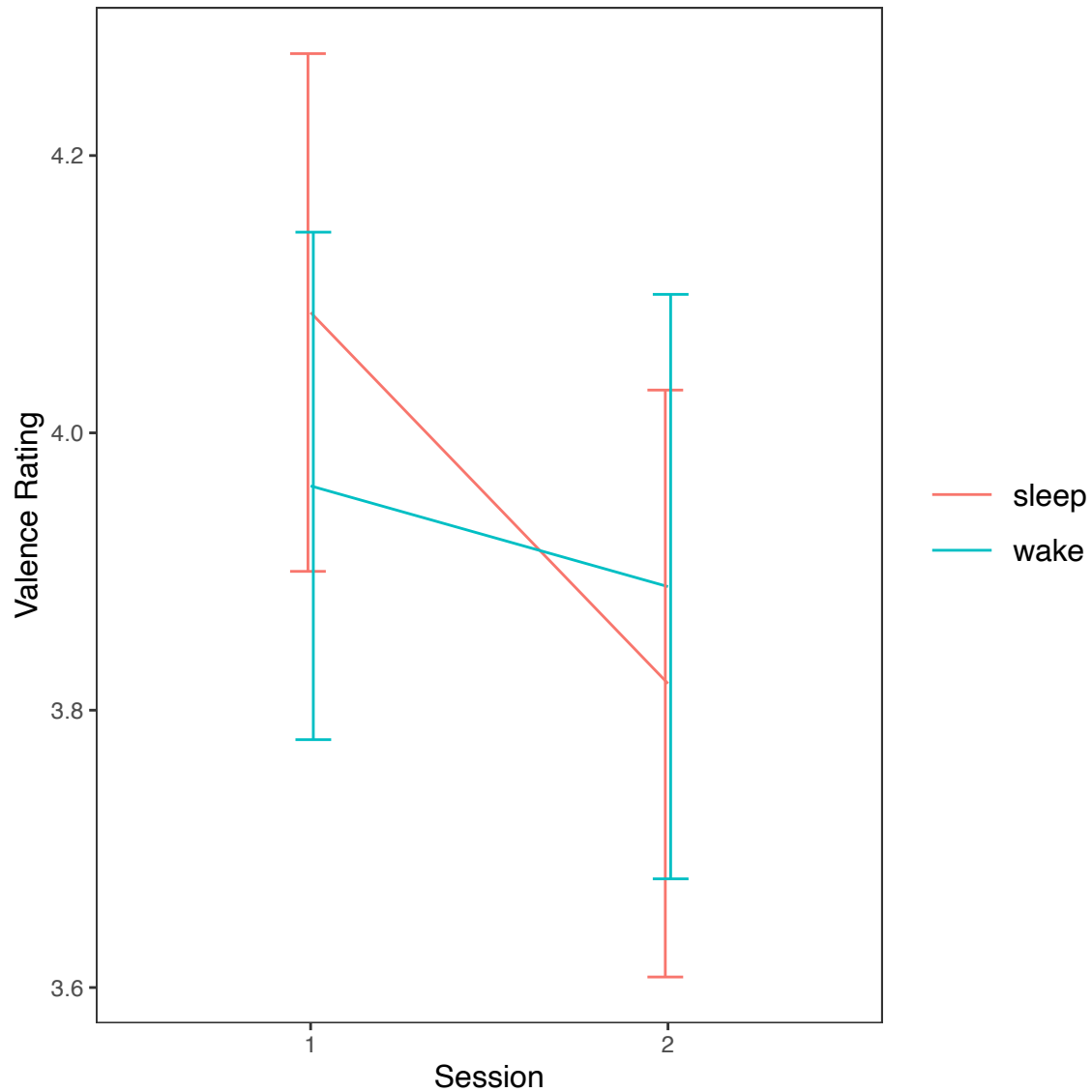


Figure 4. Mean valence ratings and 95% confidence intervals for word pairs at training (1) vs. testing (2) for the sleep and wake groups.

When arousal ratings, including both ratings at training and testing, were the outcome variable, the final model (Table 6) included significant fixed effects of congruent vs. incongruent novel metaphors, ( $\chi^2(1) = 30.59, p < .001$ ): congruent word pairs were rated as more arousing than incongruent pairs; word pair emotionality ( $\chi^2(2) = 55.594, p < .001$ ), with positive and negative word pairs rated as higher in arousal than neutral ones; valence rating at

training ( $\chi^2(1) = 131.37, p < .001$ ), with more positively valenced word pairs rated as more arousing; an interaction between time of testing and emotionality ( $\chi^2(3) = 32.958, p < .001$ ), with positive emotional pairs declining in arousal, negative emotional pairs increasing in arousal ratings at testing, and with no change for neutral word pairs (see Figure 5); and an interaction between time of testing and valence rating ( $\chi^2(1) = 56.784, p < .001$ ): the positive relationship between arousal and valence was stronger at training than at testing.

Table 6. Summary of fixed effects in final model of arousal ratings across training and testing sessions.

Predictors	Estimate	std. Error	CI	<i>t</i>	<i>p</i>
(Intercept)	0.56	0.15	2.26 – 2.85	17.18	< .001
Congruent vs. incongruent word pairs	0.28	0.05	0.18 – 0.38	5.33	< .001
Emotionality [negative vs. neutral]	0.57	0.14	0.29 – 0.85	3.97	< .001
Emotionality [positive vs. neutral]	0.79	0.15	0.50 – 1.07	5.37	< .001
Valence rating at training	0.35	0.03	0.30 – 0.40	13.82	< .001

Emotionality [neutral] ×	0.95	0.15	0.66 – 1.24	6.47	< .001
Training vs. Testing					
session					

Emotionality [negative] ×	0.69	0.10	0.49 – 0.89	6.79	< .001
Training vs. Testing					
session					

Emotionality [positive] ×	0.08	0.21	0.67 – 1.49	5.17	< .001
Training vs. Testing					
session					

Valence training ×	0.25	0.03	-0.31 – 0.18	-7.56	< .001
Training vs. Testing					
session					

---

Observations	3600, participants = 40, word pairs = 45
--------------	--

Marginal R <sup>2</sup> / Conditional R <sup>2</sup>	0.214 / 0.314
--	---------------

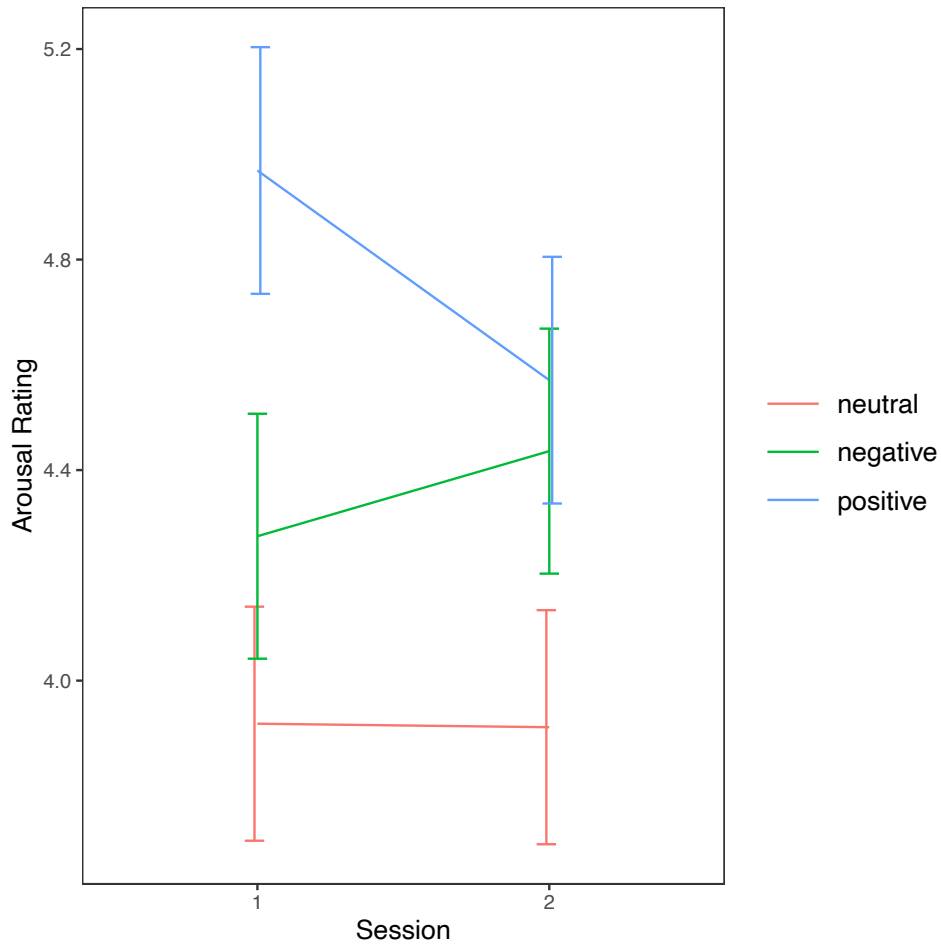


Figure 5. Mean arousal ratings and 95% confidence intervals for word pairs at training (1) vs. testing (2) for word pairs with different emotionality.

#### 4. Discussion

The present study examined the way in which figurative language is constructed, and imbued with emotional affect. We tested the role of sleep on the learning of and memory for both novel and conventional metaphorical word pairs varying in emotionality. To our knowledge, this was the first attempt at investigating memory consolidation for stimuli whose emotive content is not given, but develops in tandem with learning of the expressions

themselves. We investigated novel metaphors, including both positively and negatively valenced stimuli, and including measures of both emotional valence and arousal.

We found a significant main effect of sleep versus wake group on recognition accuracy, however no difference for recall of definitions, therefore only partially supporting the large body of research indicating a likely overall benefit of sleep to both learning and memory (Berres & Erdfelder, 2021; Rasch & Born, 2013). Also, definitions were less accurately recalled at testing than at training, showing memory decay over time.

We further predicted that sleep consolidation would particularly benefit learning and memory for more complex items (Ashton et al., 2022; Sio et al., 2013), i.e., novel metaphors, and in particular the ones with incongruent definitions. However, our incongruent items turned out to be particularly difficult for participants to generate definitions for, and remember (7% average definition accuracy); thus, sleep consolidation was unlikely to help for poorly encoded items (Wernette et al., 2024). In fact, we observed better accuracy and speed of recognition and better recall of definitions of *congruent novel* than incongruent novel metaphors. These incongruent novel metaphors are evidently difficult to learn, in particular in terms of generating definitions – though they were recognised at a rate of 84% accuracy during testing. However, investigating how such expressions might become embedded in the speaker’s mind clearly requires additional exposure and training than is present in our paradigm. Furthermore, we observed better recall of definition of *conventional* than novel metaphors. This may be due to the fact that conventional metaphorical meanings are known to participants, they don’t need to be learned, and therefore recall performance is higher than for newly-learned items. This suggests an overall advantage of a consolidation period for known compared to newly-learned items, and for new items whose definition is more intuitively sensible, compared to incongruent items. This is in line with empirical

research showing that semantically transparent figurative expressions, which are more congruent with their figurative meaning, are processed faster and therefore more easily than opaque expressions (Bell & Schäfer, 2016; Gibbs, Nayak & Cutting, 1995). This may in turn lead to better memory for transparent/congruent word pairs. However, no differences between sleep and wake were found on recall of definitions.

The lack of advantage for sleep than wake in recall of definitions could be explained by the fact that recall of definitions was tested three times, while recognition was only assessed once at testing. According to Antony et al. (2017), repeated retrieval of information may act as a fast route to memory consolidation, strengthening the links between hippocampal and neocortical representations, and leading to greater reliance on neocortical representations over time. Sleep is particularly beneficial for memory consolidation from hippocampus to neocortex, and thus no added benefit of sleep consolidation may be observed when online fast consolidation has taken place.

In addition, we found that novel metaphors were more positively rated after consolidation than at training, whereas highly familiar conventional expressions were less positively rated after consolidation than at training. In line with this, congruent novel metaphors were rated as more arousing than incongruent ones, independent of time of testing. Both of these results are in line with the Optimal Innovation Hypothesis (Giora et al., 2004), according to which highly familiar and overly repeated stimuli, i.e., conventional metaphors in our case, lead to potential disengagement and boredom (thus less positive ratings at testing than training) while optimally innovative stimuli, i.e., novel congruent metaphors in our case, elicit interest and engagement (hence, higher arousal ratings than novel incongruent metaphors), as long as they are not too difficult to process, which is the case for our novel incongruent pairs.



When it comes to effects of emotive content of our stimuli, recognition accuracy was better for increasingly arousing word pairs, whereas definitions were better recalled for increasingly more negative pairs. Furthermore, rated arousal at training was associated with higher positive valence at testing, and increasingly more positive valence at training with higher arousal at testing. This pattern seems to support an overall memory advantage for more emotive word pairs (Walker, 2021), and a relationship between positive valence and arousal. Our finding of these effects is mostly observed when analysing the more sensitive, continuous ratings provided by the same participants who took part in the study rather than by pre-determined categories of positive, negative and neutral pairs (although individual ratings essentially confirmed our categorisation); also, the effects are sometimes driven by valence and sometimes by arousal. Due to limitations in the number of stimuli we could use, to avoid excessive cognitive demand and overly long task duration, our experimental design cannot help resolve existing inconsistencies in the literature regarding the specific role of valence and arousal in memory consolidation (e.g., Adelman & Estes, 2013; Kensinger & Corkin, 2003; Reid et al., 2022; Schäfer et al., 2020); nor it can help clarify whether sleep consolidation is more beneficial for emotive stimuli than consolidation during wake.

Interestingly, the results of the current study partially support the sleep to forget, sleep to remember hypothesis (SFSR; Walker, 2021) as participants showed a reduction in their affective ratings of the word pairs. In line with the “sleep to forget” hypothesis, the sleep group rated word pairs as significantly less valenced at testing than at training, whereas the wake group showed no difference in affective ratings. In addition, further effects point toward a reduction of perceived emotive tone after a period of consolidation regardless of whether that period contained sleep or wake: 1) positive words were rated as less positive and negative words as less negative at testing than at training, while neutral words remained the same; in line with this, positive pairs declined in arousal rating after training, negative pairs

increased, but neutral pairs remained the same; 2) increasingly more arousing word pairs were rated as less positive at testing than at training, showing a reduction in strength of the relationship between arousal and positive valence; this was confirmed by the interaction between time of testing and valence ratings at training in predicting arousal ratings at testing. Even though these additional results showed no advantage of sleep versus wake, they are in line with the SFSR hypothesis which suggests the affective tone of a memory is reduced over time, *and* after sleep. Though there is variability in the effect of emotional content on sleep-based memory consolidation (Lipinska et al., 2019; Pereira et al., 2023; Schäfer et al., 2020) the contribution of our study may point to stimulus constraints where emotional effects are elicited by sleep: when emotional content is constructed at the same time as memorisation is developing. Our results point to the importance of distinguishing both valence and arousal in effects of memory (see also Reid et al., 2022).

It is important to consider whether changes in emotion from the first to second session might be explained in terms of regression to the mean. In Figure 3, we see higher valence values at session 1 becoming lower at session 2, and vice versa: lower values at session 1 becoming higher at session 2. However, whereas valence might orient to the mean, there is no similar pattern for levels of arousal. Figure 5 shows that, whereas arousal for neutral stimuli remains constant, and arousal ratings for positive stimuli reduce toward the mean for the neutral stimuli, arousal ratings for negative stimuli diverge from arousal for the neutral stimuli – increasing from session 1 to session 2.

The results of the present study require further investigation with an increased sample size and a larger set of stimuli, to further enhance power of the effects found and to be able to explore interactive effects between emotive content and type of metaphor. Our study was powered to detect effect sizes of  $d = 0.8$  or greater, which is a relatively large effect size,

however a larger sample size may enable greater nuance to be determined in future studies of learning of metaphors, and their accumulation of emotional content. As noted by Davidson et al. (2021), sleep studies tend to be relatively low in power – their meta-analysis of emotion and memory studies, for instance, showed that, for recognition memory, studies of sleep had a mean sample size of 40 (SD = 23, range: 10-84). Increasing sample size, then, can help to increase sensitivity to effects that are there, and also minimise risks of finding spurious effects (Button et al., 2013; Cordi & Rasch, 2021).

To conclude, the current study partially supports a benefit of sleep in the learning of metaphorical items, in the consolidation of emotional memories, and in the reduction of their perceived emotive tone after consolidation. However, it provides support for an overall benefit of a period of consolidation, whether after sleep or wake, in the learning of and memory for novel metaphors, more emotive word pairs, and in the reduction of their perceived emotive tone. It also extends previous findings to stimuli whose emotive content is learned, and to positive in addition to negative stimuli.

## References

- Adelman, J. S., & Estes, Z. (2013). Emotion and memory: A recognition advantage for positive and negative words independent of arousal. *Cognition*, 129(3), 530-535. <https://doi.org/>
- Antony, J. W., Ferreira, C. S., Norman, K. A., Wimber, M. (2017). Retrieval as a fast route to memory consolidation. *Trends in Cognitive Sciences*, 21, 573-576. <https://doi.org/10.1016/j.itcs.2017.05.001>
- Ashton, J. E., Staresina, B. P., & Cairney, S. A. (2022). Sleep bolsters schematically incongruent memories. *PloS One*, 17(6), e0269439. <https://doi.org/10.1371/journal.pone.0269439>
- Bell, M. J., & Schäfer, M. (2016). Modelling semantic transparency. *Morphology*, 26, 157-199. <https://doi.org/10.1007/s11525-016-9286-3>
- 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>
- Bohrn, I. C., Altmann, U., & Jacobs, A. M. (2012). Looking at the brains behind figurative language – A quantitative meta-analysis of neuroimaging studies on metaphor, idiom, and irony processing. *Neuropsychologia*, 50, 2669-2683. <https://doi.org/10.1016/j.neuropsychologia.2012.07.021>
- Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews. Neuroscience*, 14(5), 365–376. <https://doi.org/10.1038/nrn3475>
- Cameron, L. (2008). Metaphor and talk. In R.W. Gibbs (Ed.). *The Cambridge handbook of metaphor and thought*. Cambridge University Press.
- Cellini, N., Torre, J., Stegagno, L., & Sarlo, M. (2016). Sleep before and after learning promotes the consolidation of both neutral and emotional information regardless of REM presence. *Neurobiology of Learning and Memory*, 133, 136–144. <https://doi.org/10.1016/j.nlm.2016.06.015>
- Chambers, A. M., & Payne, J. D. (2014). Laugh yourself to sleep: memory consolidation for humorous information. *Experimental Brain Research*, 232, 1415-1427. <https://doi.org/>
- Citron, F. M. M., & Goldberg, A. E. (2014). Metaphorical sentences are more emotionally engaging than their literal counterparts. *Journal of Cognitive Neuroscience*, 26(11), 2585-2595. [https://doi.org/10.1162/jocn\\_a\\_00654](https://doi.org/10.1162/jocn_a_00654)
- Citron, F. M. M., Güsten, F., Michaelis, N., & Goldberg, A. E. (2016). Conventional metaphors in longer passages evoke affective brain response. *NeuroImage*, 139, 218-230. <https://doi.org/10.1016/j.neuroimage.2016.06.020>
- Citron, F. M. M., Lee, M., & Michaelis, N. (2020a). Affective and psycholinguistic norms for German conceptual metaphors (COMETA). *Behavior Research Methods*, 52, 1056-1072. <https://doi.org/10.3758/s13428-019-01300-7>
- Citron, F. M. M., Michaelis, N., & Goldberg, A. E. (2020b). Metaphorical language processing and amygdala activation in L1 and L2. *Neuropsychologia*, 140, 107381. <https://doi.org/10.1016/j.neuropsychologia.2020.107381>

- 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>
- Dastgheib, M., Kulanayagam, A., & Dringenberg, H. C. (2022). Is the role of sleep in memory consolidation overrated? *Neuroscience and Biobehavioral Reviews*, 140(104799), 104799. <https://doi.org/10.1016/j.neubiorev.2022.104799>
- Davidson, P., Jönsson, P., Carlsson, I., & Pace-Schott, E. (2021). Does sleep selectively strengthen certain memories over others based on emotion and perceived future relevance? *Nature and Science of Sleep*, 13, 1257–1306. <https://doi.org/10.2147/NSS.S286701>
- Diekelmann, S., & Born, J. (2010). The memory function of sleep. *Nature Reviews Neuroscience*, 11(2), 114–126. <https://doi.org/>
- Diekelmann, S., Landolt, H., Lahl, O., Born, J., & Wagner, U. (2008). Sleep loss produces false memories. *PloS one*, 3(10), e3512–e3512. <https://doi.org/>
- Dumay, N., & Gaskell, G. (2007). Sleep-Associated Changes in the Mental Representation of Spoken Words. *Psychological Science*, 18(1), 35–39. <https://doi.org/>
- Fainsilber, L., & Ortony, A. (1987). Metaphorical uses of language in the expression of emotions. *Metaphor and Symbolic Activity*, 2, 239–250. <https://doi.org/>
- Festl, E. (2010). Neuroimaging of text comprehension: Where are we now? *Italian Journal of Linguistics*, 22, 61–68. <https://doi.org/>
- Forgács, B., Lukács, Á., & Pléh, C. (2014). Lateralized processing of novel metaphors: Disentangling figurativeness and novelty. *Neuropsychologia*, 56, 101–109. <https://doi.org/>
- Frankland, P.W. & Bontempi, B. (2005). The organization of recent and remote memories. *Nature Reviews Neuroscience*, 6, 119–130. <https://doi.org/10.1038/nrn1607>
- Gibbs, R. W., Lima, P. L. C., & Francozo, E. (2004). Metaphor is grounded in embodied experience. *Journal of Pragmatics*, 36, 1189–1210. <https://doi.org/>
- Gibbs, R. W., Nayak, N. P., & Cutting, C. (1989). How to kick the bucket and not decompose: Analyzability and idiom processing. *Journal of Memory and Language*, 28, 576–593. [https://doi.org/10.1016/0749-596X\(89\)90014-4](https://doi.org/10.1016/0749-596X(89)90014-4)
- Giora, R., Fein, O., Kronrod, A., Elnatan, I., Shuval, N., & Zur, A. (2004). Weapons of mass distraction: Optimal innovation and pleasure ratings. *Metaphor and Symbol*, 19, 115–141. <https://doi.org/>
- Kensinger, E. A., & Corkin, S. (2003). Memory enhancement for emotional words: Are emotional words more vividly remembered than neutral words? *Memory & Cognition*, 31(8), 1169–1180. <https://doi.org/>
- Keysar, B., & Bly, B. (1995). Intuitions of the transparency of idioms: Can one keep a secret by spilling the beans? *Journal of Memory and Language*, 34, 89–109. <https://doi.org/10.1006/jmla.1995.1005>
- Lai, V. T., & Curran, T. (2013). ERP evidence for conceptual mappings and comparison processes during the comprehension of conventional and novel metaphors. *Brain and Language*, 127, 484–496. <https://doi.org/>
- Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.

- Lipinska, G., Stuart, B., Thomas, K. G. F., Baldwin, D. S., & Bolinger, E. (2019). Preferential consolidation of emotional memory during sleep: A meta-analysis. *Frontiers in Psychology, 10*, 1014. <https://doi.org/10.3389/fpsyg.2019.01014>
- Liu, P. Q., Connell, L., & Lynott, D. (2019). Ease-of-Simulation Norms for 452 Adjective Metaphors. *PsyArXiv*. <https://doi.org/10.31234/osf.io/yxgse>
- Marshall, L., & Born, J. (2007). The contribution of sleep to hippocampus-dependent memory consolidation. *Trends in Cognitive Sciences, 11*(10), 442-450. <https://doi.org/>
- Mashal, N., & Itkes, O. (2013). The effects of emotional valence on hemispheric processing of metaphoric word pairs. *Laterality: Asymmetries of Body, Brain and Cognition, 19*(5), 511-521. <https://doi.org/>
- McGaugh, J. L. (2004). The amygdala modulates the consolidation of memories of emotionally arousing experiences. *Annual Review of Neuroscience, 27*(1), 1-28. <https://doi.org/>
- Niu, X., Utayde, M. F., Sanders, K. E., Denis, D., Kensinger, E. A., & Payne, J. D. (2024). Age-related positivity effect in emotional memory consolidation from middle age to late adulthood. *Frontiers in Behavioral Neuroscience, 18*, 1342589. <https://doi.org/>
- Payne, J. D., Chambers, A. M., & Kensinger, E. A. (2012). Sleep promotes lasting changes in selective memory for emotional scenes. *Frontiers in Integrative Neuroscience, 6*, 108. <https://doi.org/>
- Payne, J. D., Stickgold, R., Swanberg, K., & Kensinger, E. A. (2008). Sleep preferentially enhances memory for emotional components of scenes. *Psychological Science, 19*(8), 781. <https://doi.org/>
- Pereira, D. R., Teixeira-Santos, A. C., Sampaio, A., & Pinheiro, A. P. (2023). Examining the effects of emotional valence and arousal on source memory: A meta-analysis of behavioral evidence. *Emotion, 23*(6), 1740-1763. <https://doi.org/>
- Reid, A., Bloxham, A., Carr, M., van Rijn, E., Basoudan, N., Tulip, C., & Blagrove, M. (2022). Effects of sleep on positive, negative and neutral valenced story and image memory. *British Journal of Psychology, 113*(3), 777-797. <https://doi.org/10.1111/bjop.12559>
- Russell, J. (2003). Core affect and the psychological construction of emotion. *Psychological Review, 110*, 145-172. <https://doi.org/>
- Schäfer, S. K., Wirth, B. E., Staginnus, M., Becker, N., Michael, T., & Sopp, M. R. (2020). Sleep's impact on emotional recognition memory: A meta-analysis of whole-night, nap, and REM sleep effects. *Sleep Medicine Reviews, 51*(101280), 101280. <https://doi.org/10.1016/j.smrv.2020.101280>
- Sio, U., Monaghan, P., & Ormerod, T. (2013). Sleep on it, but only if it is difficult: Effects of sleep on problem solving. *Memory & Cognition, 41*(2), 159-166. <https://doi.org/>
- Walker, S., Henderson, L. M., Fletcher, F. E., Knowland, V. C. P., Cairney, S. A., & Gaskell, M. G. (2019). Learning to live with interfering neighbours: the influence of time of learning and level of encoding on word learning. *Royal Society Open Science, 6*(4), 181842. <https://doi.org/>
- Wagner, U., Gais, S., & Born, J. (2001). Emotional memory formation is enhanced across sleep intervals with high amounts of rapid eye movement sleep. *Learning & Memory, 8*(2), 112-119. <https://doi.org/>

- Walker, M. P. (2021). Sleep essentialism. *Brain*, 144(3), 697-699. <https://doi.org/>
- Weidemann, C. T., & Kahana, M. J. (2016). Assessing recognition memory using confidence ratings and response times. *Royal Society Open Science*, 3(4), 150670. <https://doi.org/10.1098/rsos.150670>
- Wernette, E. M., & Fenn, K. M. (2024). Factors influencing sleep-dependent consolidation: Sleep strengthens memory based on encoding depth but not repetition. *Journal of Sleep Research*, 33(5), e14091. <https://doi.org/>
- Wilson, J. K., Baran, B., Pace-Schott, E. F., Ivry, R. B., & Spencer, R. M. C. (2012). Sleep modulates word-pair learning but not motor sequence learning in healthy older adults. *Neurobiology of Aging*, 33(5), 991-1000. <https://doi.org/>
- Xia, T., Yao, Z., Guo, X., Liu, J., Chen, D., Liu, Q., ... Hu, X. (2023). Updating memories of unwanted emotions during human sleep. *Current Biology*, 33(2), 309-320.e5. <https://doi.org/10.1016/j.cub.2022.12.004>