RaY

Research at the University of York St John

For more information please contact RaY at ray@yorksj.ac.uk

Special issue: Research report

The role of complementary learning systems in learning and consolidation in a quasi-regular domain

Jelena Mirković [a,b,*], *Lydia Vinals* [c] *and M.G. Gaskell* [b,**]

[a] *School of Psychological and Social Sciences, York St John University, York, United Kingdom*
[b] *Department of Psychology, University of York, York, United Kingdom*
[c] *Department of Psychology, Royal Holloway, University of London, Egham, United Kingdom*

ARTICLE INFO

ABSTRACT

We examine the role of off-line memory consolidation processes in the learning and retention of a new quasi-regular linguistic system similar to the English past tense. Quasi-regular systems are characterized by a dominance of systematic, regular forms (e.g., walk-walked, jump-jumped) alongside a smaller number of high frequency irregulars (e.g., sit-sat, go-went), and are found across many cognitive domains, from spelling-sound mappings to inflectional morphology to semantic cognition. Participants were trained on the novel morphological system using an artificial language paradigm, and then tested after different delays. Based on a complementary systems account of memory, we predicted that irregular forms would show stronger off-line changes due to consolidation processes. Across two experiments, participants were tested either immediately after learning, 12 h later with or without sleep, or 24 h later. Testing involved generalization of the morphological patterns to previously unseen words (both experiments) as well as recall of the trained words (Experiment 2). In generalization, participants showed 'default' regularization across a range of novel forms, as well as irregularization for previously unseen items that were similar to unique high-frequency irregular trained forms. Both patterns of performance remained stable across the delays. Generalizations involving competing tendencies to regularize and irregularize were balanced between the two immediately after learning. Crucially, at both 12-h delays the tendency to irregularize in these cases was strengthened, with further strengthening after 24 h. Consolidated knowledge of both regular and irregular trained items contributed significantly to generalization performance, with evidence of strengthening of irregular forms and weakening of regular forms. We interpret these findings in the context of a complementary systems model, and discuss how maintenance, strengthening, and forgetting of the new memories across sleep and wake can play a role in acquiring quasi-regular systems.

* *Corresponding author.* School of Psychological and Social Sciences, York St John University, York, YO31 7EX, United Kingdom.
** *Corresponding author.*
E-mail addresses: jelena.mirkovic@york.ac.uk (J. Mirković), lydia.vinals@rhul.ac.uk (L. Vinals), gareth.gaskell@york.ac.uk (M.G. Gaskell).

1.   Introduction

A recent body of literature has highlighted the role of long-term memory processes in language learning. Several studies have established that, for example, while children and infants may be successful at initial encoding of a novel word, their long-term retention is often poor (e.g., Friedrich & Friederici, 2011; Horst & Samuelson, 2008; Kucker & Samuelson, 2012; Vlach & Sandhofer, 2012; Wojcik, 2013). Vlach and Sandhofer (2012) have shown that enhancing encoding conditions can improve the otherwise poor long-term retention. Furthermore, in children and in adults a crucial role has been established for consolidation processes, including off-line maintenance, off-line strengthening, and forgetting, which play key roles in word learning and long-term retention (e.g., Brown, Weighall, Henderson, & Gaskell, 2012; Dumay & Gaskell, 2007; Dumay, 2016; Friedrich, Wilhelm, Born, & Friederici, 2015; Henderson, Weighall, Brown, & Gaskell, 2012; Tamminen & Gaskell, 2008; Werchan & Gomez, 2014; Williams & Horst, 2014).

Davis and Gaskell (2009) applied principles from the Complementary Learning Systems (CLS) model of memory (e.g., Kumaran, Hassabis, & McClelland, 2016; McClelland, 2013; McClelland, McNaughton, & O'Reilly, 1995) to account for the role of long-term memory processes in word learning. They suggested that complementary systems in the hippocampus and neocortex underpin the ability to acquire new words and integrate them with existing linguistic knowledge for long term retention. The hippocampal system is thought to be initially more strongly involved in supporting new learning, whereas neocortical networks provide a lasting basis for retention. Communication between these systems provides a means of consolidation-related changes in the retention of new linguistic material as well as the integration of the new linguistic knowledge into the long-term lexical network. Hippocampal replay of newly established memories (e.g., Skaggs & McNaughton, 1996) is thought to underpin improvements in retention specifically related to sleep.

There is now substantial evidence supporting the role of consolidation processes, and particularly sleep-related consolidation in both the retention and integration of new linguistic knowledge (e.g., Bakker, Takashima, van Hell, Janzen, & McQueen, 2014; Dumay & Gaskell, 2007; Gais, Lucas, & Born, 2006; Schreiner & Rasch, 2015, 2017; Tamminen, Payne, Stickgold, Wamsley, & Gaskell, 2010). The improved retention of new words after a period of sleep has been causally linked to hippocampal replay in studies using targeted memory replay during sleep (e.g., Batterink, Westerberg, & Paller, 2017; Schreiner & Rasch, 2015, 2017). Similarly, the integration of new knowledge with the existing lexical networks has been directly related to sleep-related memory consolidation processes (e.g., Tamminen et al., 2010). These findings are consistent with a CLS-type model of language learning, as well as other active systems consolidation models invoking sleep as a vehicle for hippocampal replay (e.g., Born & Wilhelm, 2012).

Mirković and Gaskell (2016) further examined the extent to which the CLS model applies to language learning by focusing on the encoding schemes utilized by the two systems. In particular, the hippocampal system is thought to use a sparse encoding scheme beneficial for pattern separation, whereas the neocortical system is thought to use an overlapping, distributed encoding scheme beneficial for extracting commonalities across experiences (O'Reilly & Rudy, 2001). Mirković and Gaskell (2016) hypothesized that the differences in the encoding schemes used by the two systems may be relevant for understanding initial acquisition behavior. Specifically, the hippocampal sparse encoding may be particularly beneficial if the new learning is essentially arbitrary in terms of the relationship between two domains (e.g., form-to-meaning mapping, as in vocabulary learning). However, if new learning comprises of systematic mappings it may be more amenable to direct neocortical encoding at initial acquisition due to the ability of a distributed neocortical representational scheme to capture systematic regularities. In essence, the lower level of interference between each individual mapping would allow useful learning to occur without requiring the intervention of the hippocampus for pattern separation.

If the extent of hippocampal recruitment during initial learning depends on factors such as systematicity then it is plausible that the extent of any sleep-associated benefit in retention will depend on these same factors. Thus newly learned arbitrary mappings should show strong sleep-associated benefits in retention compared with newly learned systematic mappings that have a weaker reliance on the hippocampus during encoding. Mirković and Gaskell (2016) provided an initial test of this view using an artificial language incorporating both arbitrary and systematic mappings. As predicted, the arbitrary components of the language (mappings from form to meaning, as in vocabulary learning) benefited from a retention interval including sleep as opposed to wake, whereas the more systematic aspects of the language (involving generalization of grammatical markers) showed no similar benefit.

In the current study we further explore the extent to which properties of new mappings influence the profile of initial learning and retention within a CLS framework, specifically focusing on inflectional morphology. We developed an artificial language mimicking the main properties of one of the best-studied morphological systems, the English past tense (see McClelland & Patterson, 2002; Pinker & Ullman, 2002, for reviews). This mapping displays key properties of quasi-regular systems in that there is a dominance of a systematic regular pattern (as in walk-walked, or jump-jumped), together with different degrees of deviation from that pattern in irregulars or exceptions (as in sit-sat or go-went) (see Seidenberg & Plaut, 2014, for a review). This type of quasi-regularity is found across many cognitive domains, from spelling-to-sound mappings to semantic cognition (McClelland, 2015). The study of the learning, representation, and use of these systems has benefitted from computational modeling that shares the distributed encoding aspects of the CLS (e.g., Armstrong, Dumay, Kim, & Pitt, 2017; Harm & Seidenberg, 1999, 2004; McClelland & Rogers, 2003; Mirković, Seidenberg, & Joanisse, 2011; Plaut, McClelland, Seidenberg, & Patterson, 1996; Woollams, Joanisse, & Patterson, 2009), but critically it has not considered the involvement of the hippocampal system and moreover the complementary contributions of the two systems in initial learning and long-term

retention (c.f. Schapiro, Turk-Browne, Botvinick, & Norman, 2017). Thus understanding the extent to which the CLS framework applies to learning a quasi-regular inflectional system will have important implications for learning and representation in cognition more broadly.

The existence of a systematic, regular pattern alongside irregular or exceptional items typical of quasi-regular systems allowed us to examine the specific role of the two learning systems at initial acquisition and off-line consolidation. In particular, the systematic aspects of regulars may make them more likely to make use of the distributed neocortical scheme at initial acquisition, whereas the exceptional nature of irregulars would benefit from the sparse encoding scheme and pattern separation of the hippocampal system. The difference in relative reliance on the two systems at initial acquisition might have crucial implications for consolidation: the initial stronger reliance on the hippocampal system for irregulars could make them subject to stronger consolidation effects. Thus, when assessed after a delay, one might expect a greater change in behavioral performance for irregular relative to regular forms. We explored these aspects of the CLS as applied to learning a novel morphological system by focusing on generalization behavior.

To our knowledge, this is the first study to explore the role of consolidation processes in learning a new inflectional system. Previous studies on the role of consolidation in morphological learning explored the acquisition of derivational morphology, and in particular new affixes for existing vocabulary items (e.g., Leminen et al., 2016; Tamminen, Davis, & Rastle, 2015; Tamminen, Davis, Merkx, & Rastle, 2012). For example, Tamminen and colleagues trained participants on novel derivational affixes for existing English words by teaching them, for instance, that a *sleepnule* was a person asked to sleep for a medical experiment, and a *climbnule* a person climbing dangerous mountains. Participants were tested immediately after training or after a delay (2 days or 1 week). They were tested on their ability to generalize the acquired affix-to-meaning mappings (e.g., -*nule* = a person) by combining the trained affixes with English words not presented during training. The generalization tests included tests of explicit knowledge of the affix-to-meaning mappings, and more implicit tests such as novel word naming in the context of sentences semantically congruent or incongruent with the meaning of the novel affixes. Tamminen et al. (2015) showed that in explicit tasks participants were able to generalize the meaning of the new affixes both at the immediate and the delayed test. However, in the more implicit naming tasks likely to engage existing linguistic knowledge, the benefit of learning only appeared at the delayed test when participants were faster to name the novel words in sentences congruent versus incongruent with the meaning of the novel affixes. This facilitation was absent at the immediate test.

The current study differs from previous studies of learning and consolidation of morphological systems in three ways. First, unlike previous studies that relied on existing linguistic knowledge, we used a fully artificial language with a new vocabulary and new affixes. This decision was motivated by the fact that the role of the complementary learning systems at both initial acquisition and offline consolidation is crucially influenced by the extent to which new learning is supported by existing knowledge (e.g., Himmer, Müller, Gais, & Schönauer, 2017; McClelland, 2013; Tse et al., 2007). Implementing the new inflectional system within a new vocabulary allowed us to focus specifically on new learning while minimizing the influence of existing linguistic knowledge. Second, the current study has a finer-grained focus on the initial period of consolidation, unlike previous studies that compared performance immediately after learning with delayed performance a day (Leminen et al., 2016) or a week (Tamminen et al., 2015) later. While such comparisons are valuable, they fail to assess the potentially discriminable contributions of time spent asleep versus awake on the representation of the new system (c.f. Sweegers & Talamini, 2014; Werchan & Gomez, 2014). The current study therefore compared a range of delay periods within the first 24 h in order to provide a better understanding of the processes involved in retention and consolidation of morphological learning. The third distinctive aspect of the current study is the focus on the learning and consolidation of a new inflectional system. Unlike derivational morphemes, inflectional morphemes have direct consequences for linguistic processing outside individual words; for example, in English, a plural morpheme −*s* on a subject noun crucially influences the form of the verb due to agreement (e.g., the cat *is* on the roof, *vs* the cats *are* on the roof). This property of inflectional morphology has led to models that incorporate fundamental differences in how derivational and inflectional affixes are processed (e.g., Bozic & Marslen-Wilson, 2010; Laudanna, Badecker, & Caramazza, 1992). However, alternative models describe both as simply points on a continuum, and suggest that both can be viewed as regularities in the form-to-meaning mappings (e.g., Bybee, 1995a; Hay & Baayen, 2005; Seidenberg & Gonnerman, 2000). Thus findings that relate to how both types of morphological systems are learned and supported by the CLS may have important implications for the debate on the nature of the differences between derivational and inflectional morphology.

The artificial inflectional system used in the current study mimicked several key properties of the English past tense and other inflectional systems. First, the majority of inflected forms conformed to a systematic, regular pattern (as in walk-walked), and a minority had an irregular form. Second, the irregular forms had higher token frequency (e.g., Bybee & Slobin, 1982; Bybee, 1995b). In natural languages, this property is thought to help preserve irregular forms in the language (e.g., go-went; Bybee, 1995a, 1995b). Finally, both regular and irregular forms varied in the extent to which a phonological cue was associated with the pattern (as in keep-kept, sleep-slept *vs* sit-sat, hit-hit, fit-fitted; Albright & Hayes, 2003).

We were particularly interested in what properties of this system would drive generalization performance assessed immediately after training and after a delay. First, based on previous work on learning and generalization in these systems, we predicted that immediately after learning previously unseen forms that did not have strong phonological overlap with any uninflected forms in the trained language would be regularized (a global form of generalization; cf. Xeroxed; Prasada & Pinker, 1993). Second, novel forms containing a phonological sequence that was shared with irregular uninflected forms in the trained language would also share their

irregular inflectional pattern, mimicking the 'islands of reliability' (Albright & Hayes, 2003) found in natural languages (a more local form of generalization). Third, previously unseen forms sharing phonological properties with both trained regulars *and* trained irregulars would lead to competition immediately after learning (cf. sit-sat, fit-fitted). Crucially, to the extent that the new learning can be explained within a CLS framework, and greater reliance on the hippocampal system for irregular forms, we predicted that a consolidation period would lead to the strengthening of irregular over regular trained items, enhancing the influence of the irregulars in the generalization process. We tested these predictions in two experiments, with participants in Experiment 1 tested either immediately after training or after a 24 h delay, and in Experiment 2 after a 12 h delay including sleep or wake.

2.   Experiment 1

In both experiments we used the same artificial language, with novel words (e.g., *rish*) referring to familiar objects (e.g., apple). The new inflectional system was implemented in the number domain, such that plural forms were created by adding a "suffix" to the unchanged singular form or "stem" (*rish + aff = rishaff* for 'apples'). As suggested above, the new morphological system incorporated several key properties of inflectional (and other quasi-regular) systems in natural languages.

The first property was the dominance of regulars: The plural forms of the novel words were designed such that the majority of items in the training set had a regular plural affix (e.g., -*aff*), and a minority had an irregular plural affix (e.g., -*eem*, -*esh*). Hence the regular plural affix had a higher type frequency than the irregular affixes, as they do in natural languages (Bybee & Slobin, 1982; Bybee, 1995b).

The second property involved the phonological characteristics of the "stems" and how they related to the affixes (e.g., Albright & Hayes, 2003; Plunkett & Marchman, 1991). As in English regular verbs (e.g., talk, hope, join) regular words in this artificial language had diverse phonological properties in the (uninflected) singular form (e.g., *rish*, *groll*, *heef*). This group was termed *no cue regulars*. A small subset of regulars shared a phonological cue in the singular form with a set of irregulars (e.g., -*arb* in *farb*, *yarb*), and these were termed *regular inconsistent* (see Table 1 for examples). All regular items had the same affix in the plural form (e.g., -*aff*: *rishaff*, *farbaff*, *yarbaff*).

Within the irregulars, one group of items had a phonological cue in the singular form (e.g., *isp* in *tisp*) that was uniquely associated with an irregular affix in the plural (e.g., *tispeem*). This condition is analogous to the phonological similarity in English irregular past tense forms where phonologically similar present tense forms have phonologically similar (irregular) past tenses (e.g., sing-sang, ring-rang). These items were termed *irregular consistent*, in that a phonological cue was consistently associated with a particular suffix. The second group of irregular items shared the phonological cue in the singular form with the subset of regular items described above (-*arb* items) and had their own irregular affix (e.g., -*esh*: *harbesh*). These items were labeled *irregular inconsistent*, as the phonological cue in the singular form was inconsistent in terms of its association with regular and irregular affixes (as in English sit-sat and fit-fitted) (see Table 1 for an example set).

We included the manipulation of consistency in the mapping between the phonological cue and the affix because consistency is considered a key factor influencing learning and processing in quasi-regular systems (e.g., Harm & Seidenberg, 2004; Plaut et al., 1996; Seidenberg & McClelland, 1989). For example, Seidenberg (1992) reported that English past tense forms (e.g., baked) take longer to generate if the phonological cue in the stem (-ake) is also associated with an irregular past tense form (e.g., take-took). Similarly, in spelling-sound mappings, inconsistent words such as wave (cf. have) take longer to read out loud and are more prone to errors than consistent words such as wade (e.g., Glushko, 1979; Jared, 1997, 2002; Jared, McRae, & Seidenberg, 1990; Taraban & McClelland, 1987).

The third property of the new inflectional system that mimicked morphological systems in natural languages was higher token frequency of irregular relative to regular forms (Bybee & Slobin, 1982; Bybee, 1995b). This property also allowed us to examine generalization of regular and irregular affixes in the artificial language while keeping the total number of exposures to the two types of affixes the same.

We used performance on a generalization task as a key measure of interest. In this task, participants were presented with previously unseen "stems" (uninflected singular forms) within a linguistic context that required the production of the inflected form. The task thus engaged online language production processes, and was likely to rely on the same type of knowledge where previous studies of morphological learning found consolidation effects (e.g., Tamminen et al., 2012; Tamminen et al., 2015).

We assessed the use of regular and irregular affixes in this task in three different conditions: In the *no cue* condition, the previously unseen stems did not share phonological properties with the training set (e.g., *jeech*). In the *irregular consistent* condition, the stems shared the trained phonological cue that had been associated with the unique irregular affix (e.g., *zisp*). Finally, in the *ambiguous* condition the stems shared the trained phonological cue that had been associated with both the regular and the irregular affix (e.g., *narb*).

In Experiment 1, participants' performance on the generalization test was assessed immediately after learning or after a 24-h delay. We were particularly interested in the use of regular and irregular affixes across the three conditions, and the extent to which a consolidation opportunity, with no additional training, would influence it. Overall, to the extent that the learning of the novel inflectional system is

Table 1 − Example items of the artificial language. As in natural languages, regular items had high type, and low token frequency, while irregular items had low type, and high token frequency.

	Consistency			
	No cue/Consistent		Inconsistent	
	Singular	Plural	Singular	Plural
Regularity				
Regular	rish	rishaff	farb	farbaff
Irregular	tisp	tispeem	harb	harbesh

underpinned by complementary learning systems, immediately after learning we might expect a dominance of regular affixes for the previously unseen *no cue* items, and a dominance of irregular consistent affixes for the previously unseen items with the irregular consistent phonological cue. Importantly, for the items with the ambiguous cue we may see competition between regularizations and irregularizations, with the consolidation strengthening irregularizations.

2.1. Method

2.1.1. Participants
Fifty-two students at the University of York participated in Experiment 1 for monetary remuneration or course credit after providing a written informed consent. The protocol for both Experiments 1 and 2 was approved by the Ethics Committee at the Department of Psychology, University of York. The participants in both experiments were native English speakers and had no reported hearing or learning difficulties. Four participants were excluded from all analyses in Experiment 1 due to a lack of accurate productions in one or more conditions in the generalization test (see below for more details). Participants in Experiment 1 were randomly assigned to two groups who were both trained at the same time, but were tested either immediately after learning or after a 24-h delay. Each group had a total of 24 participants.

2.1.2. Materials
2.1.2.1. TRAINING SET. There was total of 18 novel words in the training set. All 18 words were presented in their singular/ "stem" (e.g., *rish*) (Session 1 and Session 2) and plural forms (e.g., *rishaff*) (Session 2). All 18 items were pronounceable monosyllabic English pseudowords (Rastle, Harrington, & Coltheart, 2002). They were digitally recorded by a female native British English speaker in a sound attenuated booth, sampled at 44.1 KHz.

The training set was designed such that the majority of items (12 out of 18) had a regular affix in the plural form (e.g., -*aff*), and a minority (6 out of 18) had one of the two irregular affixes (e.g., -*eem*, -*esh*). Nine out of 12 regular items were labeled *no cue regulars* as they did not have any cues to the plural affix (e.g., *groll*, *rish*, *heef*). Consistency was manipulated such that three regular and three irregular items shared the rime (e.g., -*arb*: *farb*, *harb*), rendering the -*arb* cue inconsistent regarding the mapping between the phonological cue and the plural affix. These items were labeled *regular inconsistent* (e.g., *farbaff*) and *irregular inconsistent* (e.g., *harbesh*) respectively. The remaining

three irregular items had a unique phonological cue in the stem (e.g., -*isp*) which cued a unique irregular affix (e.g., -*eem*: *tispeem*). These items were labeled *irregular consistent* (see Table 1 above for example items across all conditions).

We developed six item lists to counterbalance the assignment of the affixes and the phonological cues across different consistency and regularity conditions, such that each affix and each phonological cue was paired with different conditions across the lists (see Appendix 3 for the full item set).

The novel words were paired with pictures of familiar objects (Rossion & Pourtois, 2004), with six animals, six fruits, and six artifacts, evenly distributed among different conditions. Each new word in the singular form was paired with a picture of one referent, and each plural form was paired with a picture of the same three referents (Fig. 1). There was no phonological overlap between the novel words and the existing English words for the referents.

During training, each irregular plural item was presented more often than each regular item, to capture the higher token frequency of irregular forms found in natural languages. Specifically, each irregular plural item was presented for a total of 24 times, whereas each regular plural item was presented for a total of 6 times. This manipulation allowed us to keep the total frequency of exposure to the regular and irregular affixes constant across conditions (Table 2). All items were presented the same number of times in the singular form (16 in Session 1, and 14 in Session 2).

2.1.2.2. GENERALIZATION SET. A total of 24 pseudowords were used as novel uninflected singular forms in the generalization test. They were all pronounceable monosyllabic English pseudowords (Rastle et al., 2002). Half of the generalization items did not share onsets or rimes with the training set (Appendix 3). Thus this subset of items did not have any phonological cues to the plural affixes. The other half were novel words that shared the phonological properties of the different conditions in the training set. Specifically, 6 of these items had a phonological cue that had been associated with the irregular consistent plural affix (e.g., -*isp* in *zisp*), and 6 had a phonological cue that had been associated with both a regular and an irregular plural affix in the training set (e.g., -*arb* in *narb*). Out of this total pool of items, we created two subsets to match different lists at training, with each containing a total of 18 items, with 6 items in each of the following conditions based on the cue they provided to the plural suffix: *no cue* (e.g., *jeech*), *irregular consistent* (e.g., *zisp*) and *ambiguous* between the regular and irregular inconsistent affix (e.g., *narb*). All
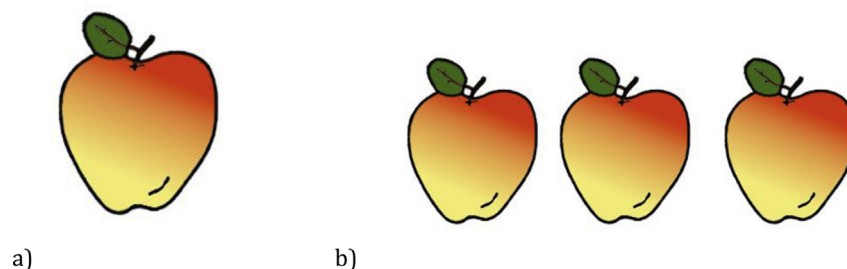

Fig. 1 – **Example picture from the training set: a) a picture that was paired with a singular form of a novel word (e.g., *rish*); b) a picture that was paired with a plural form of a novel word (e.g., *rishaff*).**

Table 2 — Frequency of the plural forms at training.

	number of items (type frequency)	item (token) frequency	affix frequency (number of items x item frequency)	total affix frequency
Regulars				
no cue	9	6	54	
inconsistent	3	6	18	72
Irregulars				
consistent	3	24	72	72
inconsistent	3	24	72	72

generalization items were digitally recorded (sampled at 44.1 KHz) in a sound attenuated booth by the same speaker as the training items.

2.1.3. Procedure

2.1.3.1. TESTING SCHEDULE. All participants were trained on the singular forms in Session 1, which occurred in the morning, lasting approximately 45 min. For Session 2, participants came back to the lab at 6 pm a week after Session 1, when they were trained on the plural forms, and continued with the exposure to the singular forms. Participants were also trained on a non-verbal declarative memory task and a procedural memory task (as these tasks were not relevant for the questions of the current study, they are presented in Appendix 4). The training part of the session lasted approximately 1 h 45 min. All participants took a 20 min break after the training phase of Session 2. Upon return to the lab half of the participants stayed on for the immediate test (0 h delay group), and half were asked to return at 8 pm on the following day (24 h delay group), when they were tested on the generalization test, and the declarative and procedural memory tests. The test phase lasted approximately 20 min.

2.1.3.2. SESSION 1. At Session 1, participants were trained on the singular forms using a word repetition task. Immediately after training, their memory for the novel words was tested using cued recall (picture naming), and 2AFC recognition. Word repetition and cued recall were implemented in DMDX (Forster & Forster, 2003), and 2AFC recognition in E-Prime (Psychology Software Inc.). Auditory stimuli were delivered via headphones.

2.1.3.2.1. WORD REPETITION. Each trial started with a fixation cross (500 msec), which was followed by the auditory presentation of the novel word; 300 msec post-word onset the corresponding picture and written word form were displayed for 4000 msec. Participants were told they were learning words of a new language, and were required to repeat each word out loud. They were told they would be tested at a later point on how well they remembered the meanings of the new words. There were two practice items at the beginning of the task. Each word was presented for a total of 16 times over 4 blocks. The order of items was randomized for each participant.

2.1.3.2.2. CUED RECALL. Each trial started with a fixation cross (500 msec), followed by the presentation of a target picture. The task was to name the picture out loud using the

novel words. There were two practice items at the beginning of the task.

2.1.3.2.3. 2AFC RECOGNITION. Each trial started with a fixation cross (500 msec), followed by the auditory presentation of the novel word and two pictures on each side of the screen. The pictures stayed on the screen for 4000 msec or until the participant made the response. Participants were required to press a button on the keyboard (1 or 9) corresponding to the picture which matched the novel word (1 for the picture on the left, 9 for the picture on the right). The two alternative pictures were always of the same semantic category. Each novel word was presented once. The position of the target picture on the screen was counterbalanced across trials.

2.1.3.3. SESSION 2. The order of training tasks in Session 2 was as follows, with a short break after task 4:

1) non-verbal declarative memory training
2) procedural memory training
3) word repetition with singular forms only
4) cued recall for singular forms
5) procedural memory training
6) word repetition with singular and plural forms
7) cued recall for singular and plural forms
8) 2AFC recognition for singular and plural forms.

Here we describe the language tasks, and we provide the description of the declarative and procedural memory tasks in Appendix 4. As in Session 1, all auditory stimuli were presented via headphones.

2.1.3.3.1. WORD REPETITION AND CUED RECALL FOR SINGULARS ONLY. Participants performed one block of the repetition task with the singular forms (the same task as in Session 1) with a total of 8 repetitions per item. This was followed by cued recall for the singular forms (the same task as in Session 1).

2.1.3.3.2. WORD REPETITION, SINGULARS AND PLURALS. Participants performed the same word repetition task as described in Session 1, but in this case it included both singular and plural forms. Singular word forms of the novel word were paired with pictures of individual objects, whereas the plural word forms were paired with pictures of three items of the same kind (Fig. 1). The total number of repetitions for each plural form varied from 6 to 24 depending on the condition (see Table 2), and each singular form was presented 6 times in

this task. The task was administered across three blocks, with a short break between the blocks.

2.1.3.3.3. CUED RECALL. Participants performed the same cued recall task as in Session 1, which in this case included pictures of both individual and multiple objects. We used the performance on this task as a measure of initial acquisition of the plural forms.

2.1.3.3.4. 2AFC, SINGULARS AND PLURALS. Participants performed the same 2AFC recognition task as in Session 1. In this block, both singular and plural trials were included. The two alternatives in the singular trials always included pictures of individual objects, and in the plural trials pictures of multiple objects. The two alternatives were always of the same semantic category. The target picture was paired with a different foil picture for the singular and the plural version of the trial. We used the performance on this task as a measure of initial acquisition of the plural forms.

2.1.3.4. TESTING. The test phase for both groups of participants included three tasks, performed in the following order: 1. nonverbal declarative memory test, 2. procedural memory test, 3. generalization test. The 0 h group was tested approximately 20 min after training, and the 24 h group was tested after a 24 h delay.

2.1.3.4.1. GENERALIZATION TEST. Each trial started with a fixation cross (500 msec), which was followed by the presentation of the phrase 'one [novel word in the singular form]' (e.g., 'one jeech') in the center of the screen, with the simultaneous auditory presentation of the singular form of the novel word over the headphones. The visual stimulus stayed on the screen for 1 sec. This was followed by a blank screen for 500 msec, which was followed by the visual presentation of the phrase 'three ?' which stayed on the screen for 4000 msec. The participant's task was to "say out loud whichever you thought was the appropriate form of the new word (…) to follow the word three." There was one practice trial, and the experimenter checked with the participant to make sure they understood the instructions. Each item from the generalization set was presented once. All responses that included an accurate production of the novel uninflected form (e.g., jeech) and one of the three affixes from the trained language were included in the analyses. Three participants were excluded in the 24 h group and one in the 0 h group because of a lack of productions meeting the above criteria in one or more of the cells of the design. The responses were coded as regularizations if they included a regular affix (e.g., -aff), as irregular consistent if they included the irregular affix that had been associated with a unique phonological cue (e.g., -eem), and as irregular inconsistent if they included the irregular affix that followed the phonological cue that had been associated with both regular and irregular forms (e.g., -esh).

2.1.3.5. DATA ANALYSES. Performance on cued recall and 2AFC recognition for plural forms were used as measures of initial acquisition. Cued recall was assessed using accuracy as a binary dependent measure. The recall data were analyzed using a mixed effects logistic regression, with the glmer function in

the lme4 package in R (Bates, Mächler, Bolker, & Walker, 2015). Recognition accuracy on the 2AFC task was high in both groups (M_{0h} = .99, M_{24h} = .97), so the analyses below focus on reaction times (RTs) for correct responses. The RT data were analyzed using a mixed effects linear regression, using the lmer function in the lme4 package in R (Bates et al., 2015), with logarithmically transformed RTs to address non-normality. In both tasks, group (0 h, 24 h), regularity (regular, irregular), and consistency (no cue/consistent, inconsistent) were included as effect-coded fixed factors, and a maximal structure for random effects for participants and items that allowed the models to converge (Barr, Levy, Scheepers, & Tily, 2013). For all models, we followed the maximal inclusion of intercepts and slopes in the initial model, which was reduced in a stepwise manner starting from the highest-order slopes until a model that converged was identified. The final random effects structure for all models presented in the Results section is provided in Appendix 1.

The performance on the generalization task was assessed using loglinear analysis that tested the distribution of the three types of responses (regularizations, and the two types of irregular affixes (irregular consistent, irregular inconsistent)), across the two groups (0 h, 24 h), and three different phonological cues (no cue, irregular consistent, ambiguous). The analyses were run using the loglm function in R.

In all analyses for both experiments the acceptable level of Type I error was set at .05.

2.2. Results

2.2.1. Initial acquisition

Participants' initial acquisition of the training set was assessed using cued recall and the 2AFC recognition task. Our aim was to determine whether the groups were well matched at the initial learning of the plural system, and to examine whether, across groups, the structure of the system would impact on the initial levels of learning as it does in natural languages (Albright & Hayes, 2003; Marchman, 1997; Plunkett & Marchman, 1993; Prasada & Pinker, 1993).

As assessed by accuracy in cued recall, the two groups showed the same level of learning of the plural forms (group: ß = −.24, SE = .45, z = −.53, p = .59). Across both groups, regular and irregular forms were recalled equally well (regularity: ß = .25, SE = .20, z = 1.24, p = .21), but more plural forms were recalled accurately in the no cue/consistent condition (rishaff, tispeem) relative to inconsistent plurals (farbaff, harbesh) (ß = −.50, SE = .20, z = −2.45, p = .014). There was also a significant regularity × consistency interaction (ß = 1.30, SE = .41, z = 3.19, p = .001; see Fig. 2), indicating that at initial acquisition the inconsistency in the phonological cue had a greater cost for the lower token frequency regulars than for the irregulars. There were no other significant effects (Appendix 1, Table 1).

The overall poorer performance with inconsistent plurals was confirmed in the recognition task (Fig. 3): Participants in both groups were slower to recognize inconsistent relative to consistent plural forms (ß = .08, SE = .03, t = 3.08, p = .002). There were no other significant effects (Appendix 1, Table 2).

In summary, the two groups of participants were well-matched in terms of initial acquisition on the trained plural
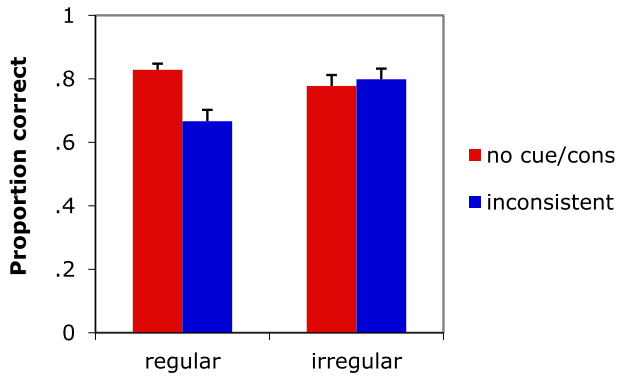
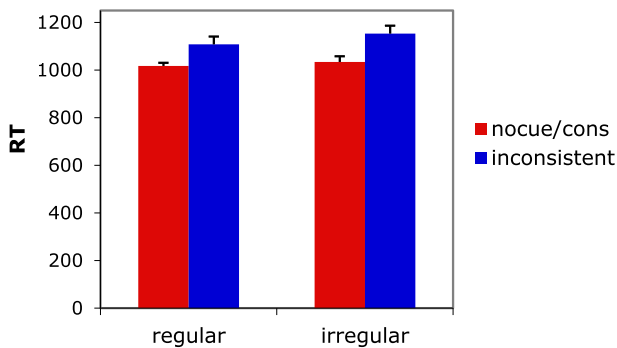**Fig. 2 — Cued recall accuracy with plural forms at the end of training.**

Fig. 3 — Recognition times on the 2AFC task for the plural forms at the end of training.

forms. There were no overall differences between the regular and the irregular items, suggesting that in both groups regular and irregular affixes were learned equally well. Both recall accuracy and recognition times were influenced by the consistency between the phonological cue and the plural affix, similar to natural languages (e.g., Albright & Hayes, 2003; Marchman, 1997; Seidenberg, 1992). In both tasks, and both groups, there was evidence of poorer initial learning for the plural forms where the phonological cue to

the affix was associated with two possible affixes (inconsistent forms). In the recall task there was also evidence that this was particularly the case for the regular/low token frequency forms.

2.2.2. Generalization

The key question we were interested in was what properties of the novel morphological system would drive performance on new, previously unseen forms when participants were required to generalize their existing knowledge of the trained forms. In addition, we wanted to examine whether a consolidation period might influence performance and specifically the relative contribution of the systematic, regular forms versus the irregular forms.

The analyses below included three possible types of responses to previously unseen novel "stems" (uninflected singular forms): *regularizations* (the production of the regular affix), and two types of *irregularizations* — the production of the irregular affix that had consistently followed a unique phonological cue, termed an *irregular consistent* response, or the production of the irregular affix that had followed the phonological cue associated with both the regular and the irregular forms, an *irregular inconsistent* response. We were interested in the frequency distribution of these types of responses across the three phonological cues in the group tested immediately versus the group tested after a 24 h delay.

The loglinear analysis showed that the distribution of responses was influenced by a three-way interaction: the difference between the saturated model and the model that excluded the group x stem cue x response type interaction was significant [χ^2 (4) = 30.21, $p < .001$]. As illustrated in Fig. 4, both groups showed a similar pattern of responses to two of the three cue types: When there was no phonological cue in the "stem", participants tended to provide a regularization response. Conversely, when there was a cue in the novel "stem" that matched the trained consistent irregulars then the corresponding irregular affix tended to be chosen. Crucially, when the phonological cue was ambiguous (matching trained regular *and* irregular forms) the type of response depended on when the participants were tested: participants tested immediately after training produced a similar number of regular and irregular inconsistent affixes, whereas participants tested with a 24-h delay predominantly produced an irregular inconsistent affix.
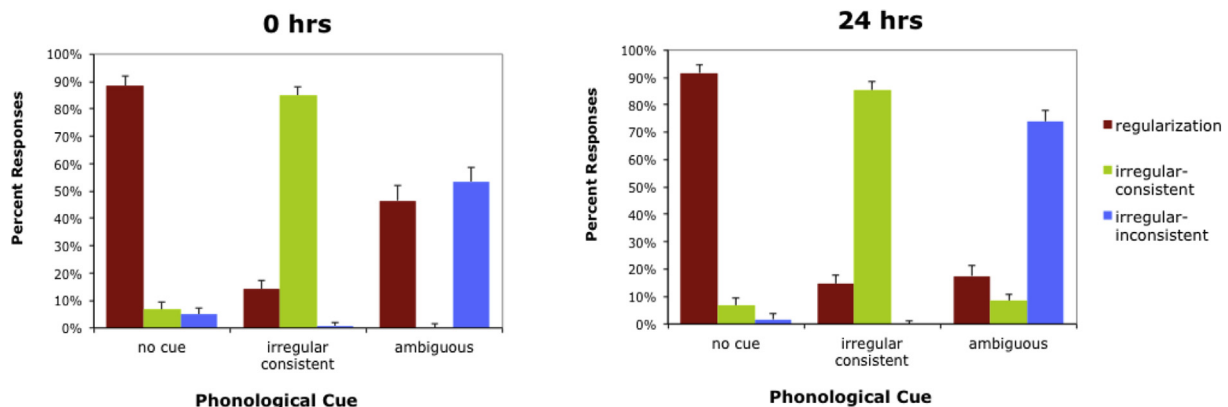

Fig. 4 — Generalization performance for the participants tested immediately (0 h) or after a delay (24 h).

In summary, generalization responses indicated that after learning a novel quasi-regular inflectional system, participants showed both regularizations and irregularizations as in natural language learning. Crucially, after a delay and with no additional training and when encountering an ambiguous phonological cue irregularizations dominated over regularization responses.

2.3.   Discussion

In this study, we trained participants on a novel morphological system that incorporated a large set of regular forms, and a smaller set of exceptions, thus mimicking other quasi-regular systems found in natural languages. At initial acquisition, and as in natural language, participants' performance was influenced by the consistency of the phonological cues to the affix, with poorer initial learning of the forms with inconsistent phonological cues. Importantly, there was no overall effect of regularity, suggesting that regular and irregular affixes were learned equally well.

Our main questions related to how participants would generalize their newly acquired knowledge of this artificial morphological system. Would they show evidence of a regular "default" inflection that could be applied to a range of forms that had no particular similarity to any training item (e.g., Prasada & Pinker, 1993)? Moreover, would they show evidence of irregularization in cases where novel uninflected forms were phonologically similar to trained exceptions? Most crucially, would the balance between these competing tendencies change over the course of 24 h after initial learning due to consolidation processes?

We found that participants produced regular affixes for previously unseen items where there was no phonological overlap with any items in the training set. This is analogous to the performance in natural languages (e.g., Xeroxed) and suggests that participants had acquired a default as a consequence of the structure of the training materials. In contrast, when the new items contained a phonological cue uniquely associated with an irregular affix, that affix dominated the responses, leading to irregular inflections. Again, this is similar to generalization performance in natural languages which shows sensitivity to subtle phonological regularities (Albright & Hayes, 2003; Marchman, 1997).

Both the default generalization of the regular plural and the more specific irregularization for items highly similar to consistent irregulars were observed very strongly 20 min after learning (observed in more than 80% of the associated trials), and these behaviors remained dominant after a delay of 24 h. This lack of change over time might be an indication that consolidation processes are of little relevance to the retention of inflectional learning. However, it is also feasible that the strong dominance of one type of generalization response for the no cue and irregular consistent items makes these conditions less sensitive to consolidation—a kind of ceiling effect. Consistent with the latter interpretation, for new items with an ambiguous phonological cue that could lead to either regularization or irregularization we found that generalization performance was well balanced between these two options soon after learning (54% irregular responses, compared to 46% regulars). This state of equilibrium altered substantially after

24 h with irregularizations now much more likely and the proportion of regular plural responses dropping to just 18%.

The change of generalization performance over the delay and without any additional training suggests that, as in vocabulary learning (e.g., Bakker, Takashima, van Hell, Janzen, & McQueen, 2015; Bakker et al., 2014; Dumay & Gaskell, 2007; Gaskell & Dumay, 2003), consolidation processes play a role in learning a quasi-regular inflectional system. The specific pattern of change suggests that the influence of exception/irregular items was enhanced over the delay, such that for the phonologically ambiguous cue irregularizations dominated generalization responses at the delayed but not the immediate test. Based on the CLS model and previous work on the role of the hippocampus in pattern separation, this may have been due to the initially stronger reliance on the hippocampal system for irregular items, and sleep-related memory consolidation benefits for hippocampally encoded memories (e.g., Tamminen et al., 2010). These processes may have been facilitated further by stronger forgetting of systematic, regular aspects of the new morphological system over wake. That is, the more systematic mappings initially more strongly reliant on the neocortical system may have been subject to stronger interference-based forgetting occurring during the wake period (e.g., Sadeh, Ozubko, Winocur, & Moscovitch, 2014). To assess the contribution of these consolidation processes to learning and retention of a quasi-regular inflectional system, in Experiment 2 we trained participants on the same new language as in Experiment 1, and we tested them after a 12 h delay which included sleep or wake.

3.   Experiment 2

In Experiment 2 we further explored consolidation processes that contributed to the change in generalization performance across the 24 h delay. We specifically focused on the contribution of sleep and wake-related consolidation processes. A substantial body of evidence has established that sleep-related memory consolidation helps strengthen new linguistic knowledge (e.g., Dumay & Gaskell, 2007; Gais et al., 2006; Schreiner & Rasch, 2015, 2017; Tamminen et al., 2010) through the process of hippocampal replay (Batterink et al., 2017; Schreiner & Rasch, 2015, 2017). In the current study, to the extent that irregular forms are more strongly encoded by the hippocampal system, we may expect greater sleep-related strengthening of irregular relative to regular forms.

During wake, although consolidation-related strengthening of new memories may occur (e.g., Dewar, Alber, Butler, Cowan, & Della Sala, 2012; Dewar, Alber, Cowan, & Della Sala, 2014), we generally see stronger influence of forgetting (see Diekelmann & Born, 2010 for a review). Few studies have explored the specific role of forgetting in language learning. One exception is the study by Werchan and Gomez (2014) who investigated the role of forgetting in word learning in 2.5-year old toddlers. They showed that wake-related forgetting of specific item-level information was beneficial for generalizations based on systematic aspects of new form-meaning mappings. This finding suggests that in addition to sleep-related strengthening, wake-related forgetting may also play a role in learning and consolidation of a new inflectional system.

To assess the influence of sleep and wake on learning and retention in a quasi-regular domain, in Experiment 2 participants were trained in the morning or in the evening and tested after a delay of 12 h which included wake or sleep. The study protocol followed Experiment 1 closely, and our primary goal was to determine the influence of these two time periods on the generalization of the new learning to previously unseen materials. To provide an additional source of evidence of the memory for the inflectional system, we also tested performance on the trained items themselves, both immediately after learning and after a 12 h delay including sleep or wake. We used an item fate analysis (Dumay, 2016; Schreiner & Rasch, 2016) to explore the extent to which systematic regular and exceptional irregular items are strengthened or forgotten over the delay period.

3.1. Method

3.1.1. Participants
Sixty-eight students at the University of York participated in Experiment 2 for monetary remuneration or course credit after providing a written informed consent. After exclusion of 8 participants due to a lack of any accurate productions in one or more conditions in the generalization test, 31 participants were included in the analyses in the wake group, and 29 in the sleep group.

3.1.2. Materials
The same training and generalization materials were used as in Experiment 1.

3.1.3. Procedure
3.1.3.1. TESTING SCHEDULE. As in Experiment 1, all participants were trained on the singular forms in Session 1, which occurred in the morning, lasting approximately 45 min. For Session 2, participants came back to the lab a week after Session 1, with the wake group starting Session 2 at 8.30 am, and the sleep group starting at 8.30 pm. During this session participants were trained on the plural forms and tested on their initial acquisition. Session 2 lasted approximately 1 h 45 min. All participants returned to the lab for the final test session approximately 12 h later (8.30 pm for the wake group,[1] 8.30 am for the sleep group), when they were tested on the generalization test, as well as on the trained items. The test phase lasted approximately 30 min.

3.1.3.2. SESSIONS 1 AND 2. The procedure was the same as in Experiment 1, aside from the following: At the beginning of Session 2 participants filled out questionnaires about their sleep habits and sleepiness [Epworth and Stanford Sleepiness Scales (Hoddes, Zarcone, Smythe, Phillips, & Dement, 1973; Johns, 1991)]. In addition, we also administered a simple reaction time task as a test of alertness (Reid, 2013).

3.1.3.3. TESTING. The test phase for both groups of participants included the same three tasks as in Experiment 1, performed in the same order. These were followed by two tasks assessing performance on the trained items (as in Session 2): cued recall for singular and plural forms, and 2AFC recognition for singular and plural forms. The test session started with the Stanford Sleepiness Scale and the alertness task (the data analyses for these tasks are presented in Appendix 5).

3.1.3.4. DATA ANALYSES. The analyses of initial acquisition and generalization were performed as in Experiment 1. For the generalization analyses we additionally compared performance across the two experiments. For this analysis, we used planned orthogonal Helmert contrasts and mixed effects logistic regression specifically to compare irregular inconsistent responses for the ambiguous stem cue across the four delay periods. The first set of contrasts compared the 24 h group (the delayed group in Experiment 1, coded as 3) with the other three groups [12 h wake, 12 h sleep, immediate (0 h in Experiment 1), each coded as −1]; the second compared the immediate group (0 h group in Experiment 1, coded as 2) with the two 12 h groups (12 h wake, 12 h sleep, each coded as −1), and the third compared the two 12 h groups [wake (1) vs sleep (−1)]. For this analysis, we focused on the ambiguous cue condition. Each generalization response for the items in this condition was coded as 1 if it included the irregular inconsistent affix, and as 0 if it was a regularization. Thus this analysis focused on the extent of irregularization at different delays. A small proportion of responses in this condition (2.8%) included an irregular consistent affix (which for the ambiguous condition represented an inappropriate inflection), and these were excluded from the analyses. The final models that converged included random intercepts by participants and by items for all contrasts, and random slopes for the first and the second set of contrasts.

For the analyses of the change of performance on the trained items over the two 12 h delays in Experiment 2, we used several measures. First, as for initial acquisition, we analyzed accuracy at cued recall, and RTs for correct responses for the 2AFC task, with session (immediate, delayed), group (wake, sleep), regularity (regular, irregular), and consistency (consistent, inconsistent) as effect-coded fixed factors. All fixed and random effects included in the models are provided in Appendix 2.

In addition, to identify properties of the new language that influenced different consolidation processes, we analyzed performance at cued recall using two measures of item 'fate', explained further in the Results section.

3.2. Results and discussion

3.2.1. Initial acquisition
As in Experiment 1, the level of initial learning of the plural forms was assessed using cued recall and 2AFC recognition. There was a similar level of initial learning in cued recall as in Experiment 1 (Fig. 5), with no difference between the wake and the sleep group (ß = .05, SE = .41, z = .12, p = .90). Participants in both groups recalled significantly more plurals in the no cue/consistent relative to the inconsistent condition (ß = −.49, SE = .18, z = −2.73, p = .006), with a significant

---

[1] Four participants reported having taken a nap during the delay period, with the average duration of 26.5 min. The results of the analyses remain the same regardless of whether these participants are included in the analyses, suggesting a negligible influence of these short sleep periods on test performance.
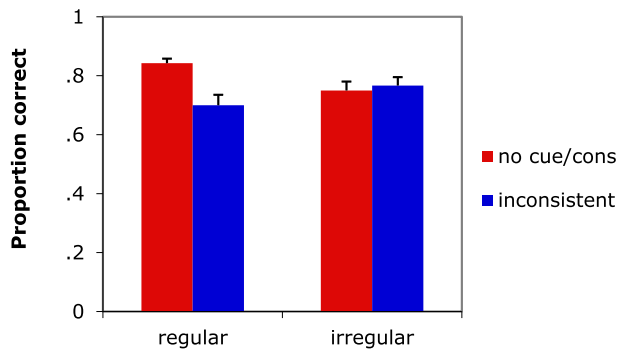
Fig. 5 — Cued recall accuracy with plural forms at the end of training.

regularity × consistency interaction (ß = 1.24, SE = .36, z = 3.41, p = .001). As illustrated in Fig. 5, the increased difficulty with forms with inconsistent affixes was problematic for the regulars, but less so for the irregulars. There were no other significant effects (Appendix 2, Table 1). This pattern of findings replicates the performance at cued recall for initial acquisition of the new morphological system found in Experiment 1.

The overall poorer performance with inconsistent items was confirmed in the recognition task (Fig. 6): participants in both groups were slower to recognize inconsistent relative to consistent plurals (ß = .11, SE = .03, t = 4.41, p < .001). There were no other significant effects (Appendix 2, Table 2).

In sum, the analyses of initial acquisition replicate the findings of Experiment 1, and overall suggest a good level of learning of the regular and the irregular suffixes, with no differences between groups. The learning of the plural forms was influenced by the consistency of the phonological cue in the stem, with poorer learning of inconsistent plurals, and in particular poorer recall of regular inconsistent forms in both groups.

3.2.2. Generalization
A key question in the current experiment was how generalization performance on the delayed test would be influenced by whether the delay included wake or sleep. In particular, we were interested in how the dominance of regular versus irregular based generalizations would change over the 12 h delay period.
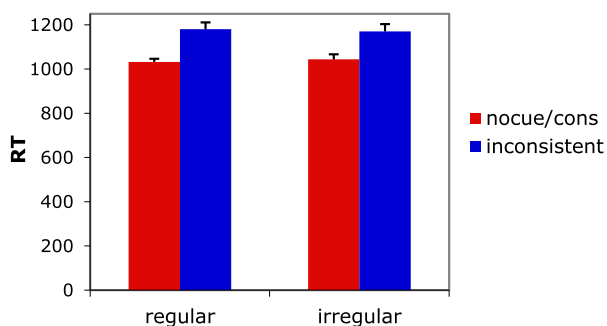

Fig. 6 — Recognition times at the 2AFC task with plural forms at the end of training.

The loglinear analysis showed that, in contrast to Experiment 1, the two groups had a similar pattern of generalization (Fig. 7): the three-way interaction between group, stem cue, and type of response did not significantly contribute to the model [χ^2 (4) = 1.86, p = .76]. As in Experiment 1, the generalization responses varied with cue type: Further model comparisons showed a significant contribution of the phonological cue x response type interaction [χ^2 (4) = 982.61, p < .0001], and no contributions from either interaction involving group [group x phonological cue: χ^2 (2) = .98, p = .61; group x response type: χ^2 (2) = 1.61, p = .45]. As illustrated in Fig. 7, when there was no phonological cue in the novel "stem" participants in both groups had a strong tendency to produce regularization responses. Conversely, when the phonological cue overlapped with the irregular consistent forms in the training set the dominant response in both groups was the irregular consistent affix. Notably, and unlike the immediate testing group in Experiment 1, in both delayed groups in the current experiment irregular inconsistent affixes constituted approximately two thirds of responses (65% in both groups) for the ambiguous cue. This difference suggests an increase in this type of generalization response across the 12-h delay relative to the immediate testing, that did not seem to vary depending on whether the delay included wake or sleep.

These analyses reveal a similar pattern of responses for two out of three types of cues as in Experiment 1, with regularizations dominating the no cue items, and irregular consistent responses dominant for the items with a phonological cue uniquely associated with an irregular affix. This suggests that the factors that underlie performance in these two conditions soon after learning remain stable across 12 h and 24 h periods regardless of whether sleep or wake intervene.

The ambiguous cue condition was the only one in Experiment 1 to show evidence of a close competition between two types of response, and was also the only one to reveal a change in performance across 24 h. Interestingly, for this cue both groups tested after a 12 h delay showed an apparent increase of irregular inconsistent responses relative to the group tested immediately after training in Experiment 1. To assess further the extent to which preference for irregular inconsistent responses emerges and increases over a delay period, we analyzed generalizations of all four groups of participants across the two experiments (tested immediately after learning, or after a 12 h wake, or a 12 h sleep, or a 24 h delay), focusing on the irregular inconsistent responses in the ambiguous cue condition. As illustrated in Fig. 8, these analyses showed that participants who were tested after 24 h had the greatest tendency to produce irregular inconsistent generalizations relative to the other three groups (24 h vs 0 h/12 h wake/12 h sleep contrast: ß = .34, SE = .09, z = 3.84, p = .0001). Participants in both 12 h groups also had a greater tendency to produce this type of response compared to the group tested immediately after training (ß = −.22, SE = .10, z = −2.16, p = .031). As in the main analyses, there were no differences between the two 12 h groups (ß = −.01, SE = .16, z = −.07, p = .94).

These findings indicate that when presented with a phonological cue that could lead to regularizations or irregularizations, when tested immediately after training
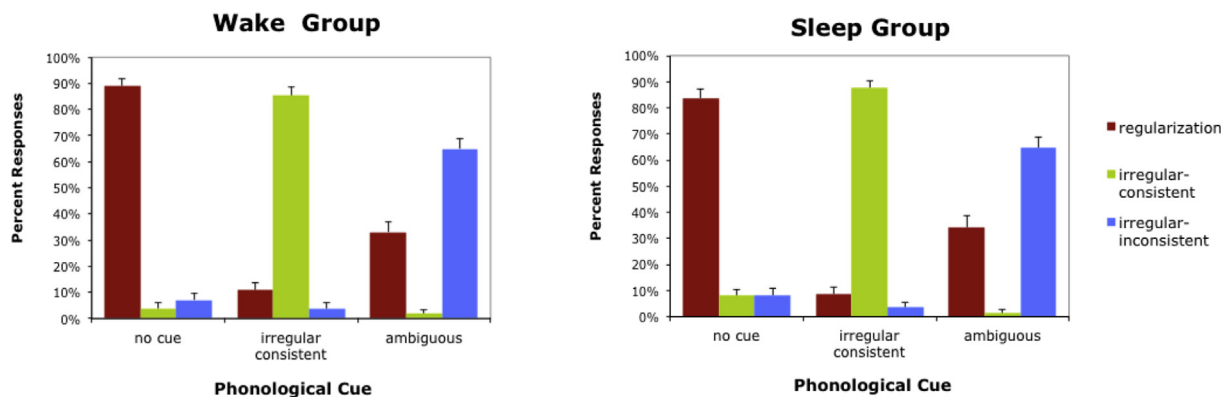
Fig. 7 − Generalization performance for the participants tested after a delay of 12 h including wake or sleep.
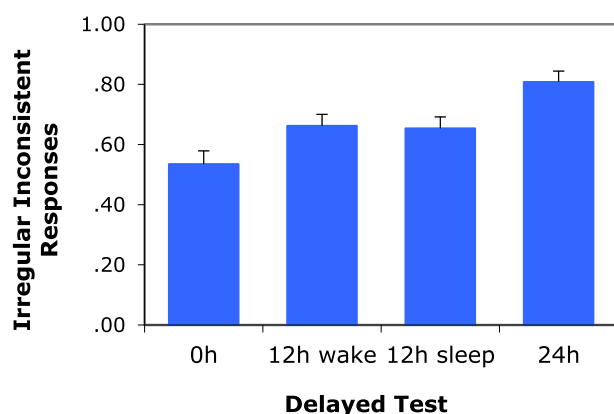

Fig. 8 − Proportion of irregular inconsistent responses in the generalization task for the ambiguous cue across the four delays.

participants showed a balanced preference for the two types of responses. However, after a 12-h delay with or without sleep there was a clear preference for irregular inconsistent responses, which increased further with an additional 12-h delay.

3.2.3. Change of memory for the trained items across the delay

The pattern of generalization performance described above may result from a mixture of maintenance, strengthening, and forgetting of different aspects of the newly learned inflectional system across sleep and wake periods. To assess the contribution of these processes, we examined the change in performance across the 12-h delay for the trained items.

We first analyzed overall accuracy at cued recall including session (immediate, delayed) as a fixed factor. This analysis confirmed that across sessions and groups participants had poorer recall of inconsistent relative to consistent items (ß = −.40, SE = .12, z = −3.47, p = .001), and this was particularly the case for regular items (regularity x consistency: ß = 1.07, SE = .23, z = 4.65, p < .001; Appendix 2, Fig. 1). Turning to the change in performance across the delay (i.e., interactions with session; Fig. 9), the wake group showed overall

poorer performance driven by poorer retention of the new plural forms over the delay relative to the sleep group (group: ß = .57, SE = .12, z = 4.93, p < .001; group x session: ß = 1.00, SE = .23, z = 4.35, p < .001; Appendix 2, Fig. 1). Regular forms were overall less well retained than irregular forms (session × regularity: ß = .50, SE = .23, z = 2.20, p = .028). There were no other significant effects (Appendix 2, Table 3).

The effects of session on cued recall performance likely reflect some combination of forgetting, maintenance or strengthening of memories in wake and in sleep. To tease these processes apart, we ran an item "fate" analysis following Dumay (2016; see also Dumay, 2018; Schreiner & Rasch, 2016). This type of analysis partitions the data into two complementary sets of trials based on whether the item was correctly recalled or not prior to the 12-h delay. For items that were not correctly recalled in the test immediately after learning, the two possible outcomes after the delay are a further inaccurate recall (described as *never learned*, coded as 0), or a correct recall (described as a *gain*, coded as 1). For items that were correctly recalled prior to the delay, the two possible outcomes are correct recall again (*maintained*, coded as 0) or inaccurate recall (*forgotten*, coded as 1). We ran two complementary analyses on these datasets using a mixed effects logistic regression with gains versus never learned and forgotten versus maintained as outcome measures, with a maximal structure of random effects for participants and items that allowed the model to converge. These analyses included group (wake, sleep), regularity (regular, irregular), and consistency (consistent, inconsistent) as effect-coded fixed factors (all effects are provided in Appendix 2, Tables 4 and 5).

For the gains versus never learned analysis, we expected to see better overall performance in the sleep group on this measure due to sleep-related strengthening of newly acquired memories. Indeed, that is what we found: as illustrated in Fig. 10, participants in the sleep group showed greater gains over the delay than the participants in the wake group (ß = 1.59, SE = .56, z = 2.85, p = .004). Overall, there were more gains for irregular forms (ß = 1.03, SE = .40, z = 2.58, p = .010). The analysis also yielded a regularity × consistency interaction, with fewest gains for regular inconsistent items in both groups (regularity x consistency: ß = 2.14, SE = .82, z = 2.59, p = .010)
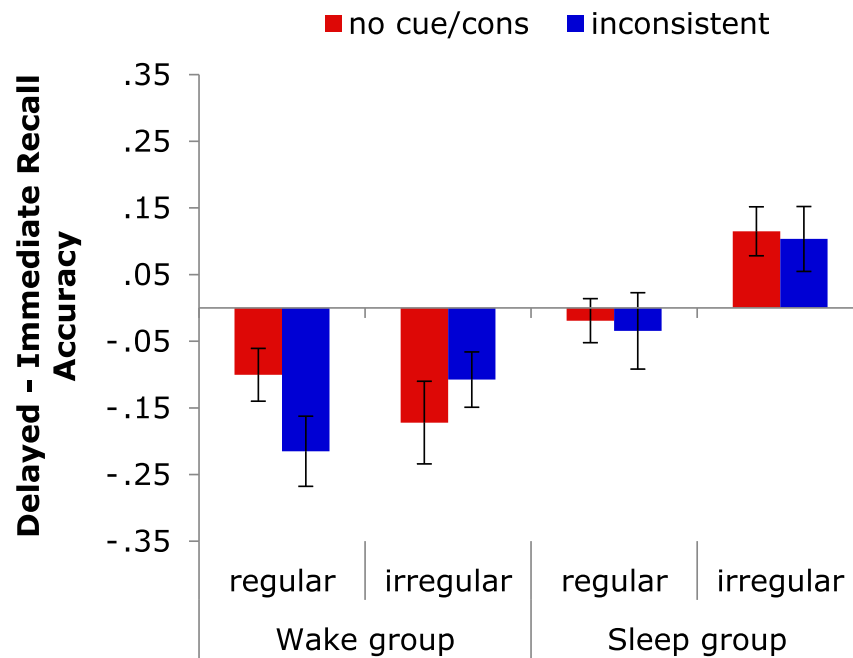
Fig. 9 — Cued recall accuracy with trained plurals illustrated using differences scores (delayed − immediate) with trained plurals for the two groups.

Fig. 10 — Proportion of items that were gained across the 12-h delay for the two groups at cued recall. The proportions are calculated out of all items that were not recalled at the immediate session, but were either gained (red and blue bars) or were not recalled at either session (never learned; grey bars).

(see also Appendix 2, Table 4). In sum, the analysis of gains showed the expected benefit of sleep, and provided evidence for the overall strengthening of irregular items, with smallest gains for regular inconsistent forms in both groups.

Turning to the analysis of items that were initially recalled correctly, we analyzed the impact of group, regularity and consistency on the likelihood of forgetting. As expected, and illustrated in Fig. 11, there was an overall greater degree of

forgetting in the wake than the sleep group (ß = −1.65, SE = .52, z = −3.20, p = .001). Overall, more inconsistent than consistent forms were forgotten (ß = .66, SE = .30, z = 2.21, p = .027). Interestingly, the pattern of forgetting in the two groups varied for different types of items: in the wake group forgetting was strongest for regular inconsistent and irregular consistent forms, unlike the sleep group (group × regularity: ß = −1.36, SE = .60, z = −2.27, p = .023; group × consistency: ß = −1.19, SE = .60, z = 2.00, p = .046; group × regularity × consistency: ß = 3.07, SE = 1.31, z = 2.34, p = .02). There were no other significant effects (Appendix 2, Table 5). This finding may suggest that the items with greatest forgetting in the wake group relied more strongly on the neocortical system at initial encoding, and thus they were subject to stronger effects of wake-activity related interference-based forgetting.

In sum, the analysis of performance on the trained items suggested that despite the very similar patterns of generalization behavior for sleep and wake participants, the underlying memory for the trained materials in these two groups was somewhat different. There was a clear difference in the overall cued recall performance between the two groups (Fig. 9), with a greater decay in performance for the wake group. This overall difference is likely a result of greater strengthening of initially weak memories in the sleep group (as evidenced by the analysis of gains), and heightened forgetting of initially strong memories in the wake group (as demonstrated by the analysis of forgetting). In addition to these general group differences, the gains and forgetting analyses revealed a more complex pattern, with the level of performance depending on consistency and regularity. The forgetting analysis revealed a particularly intricate pattern, with the group × regularity interaction showing a more equal

pattern of forgetting for regulars and irregulars in the wake, but less forgetting of irregulars than regulars for the sleep group. A group by consistency interaction and a three-way interaction between these variables showed that consistency was also important in the degree to which items were forgotten, with the wake but not the sleep group showing substantial forgetting of inconsistent regulars and consistent irregulars.

The analyses also demonstrated a general effect of regularity for the memory of the trained items across both groups. In the overall analysis and the gains analysis (and with a nonsignificant (p = .061) effect in the forgetting analysis) performance deteriorated less for the irregular items than the regulars. Putting this together with the overall benefit for sleep over wake, this result suggests that the interval between immediate and delayed testing benefits irregulars more than regulars, but that the benefit is dominated by memory gains for irregulars in sleep and by enhanced forgetting of regulars, and particularly regular inconsistent items, in wake.

The final analysis of the change of performance on the trained items over the delay provides additional evidence for overall improved performance in the sleep group on the trained items over the delay. In the 2AFC task, and as illustrated in Fig. 12, participants who slept between training and test were faster to recognize the new plural forms at the delayed test whereas there was no overall change in performance for the group who stayed awake (session: ß = −.04, SE = .02, t = −2.83, p = .005; group x session: ß = −.09, SE = .03, t = −2.99, p = .003). The overall greater difficulty with inconsistent relative to the consistent items was replicated in this analysis with participants in both groups overall slower to recognize inconsistent relative to consistent forms
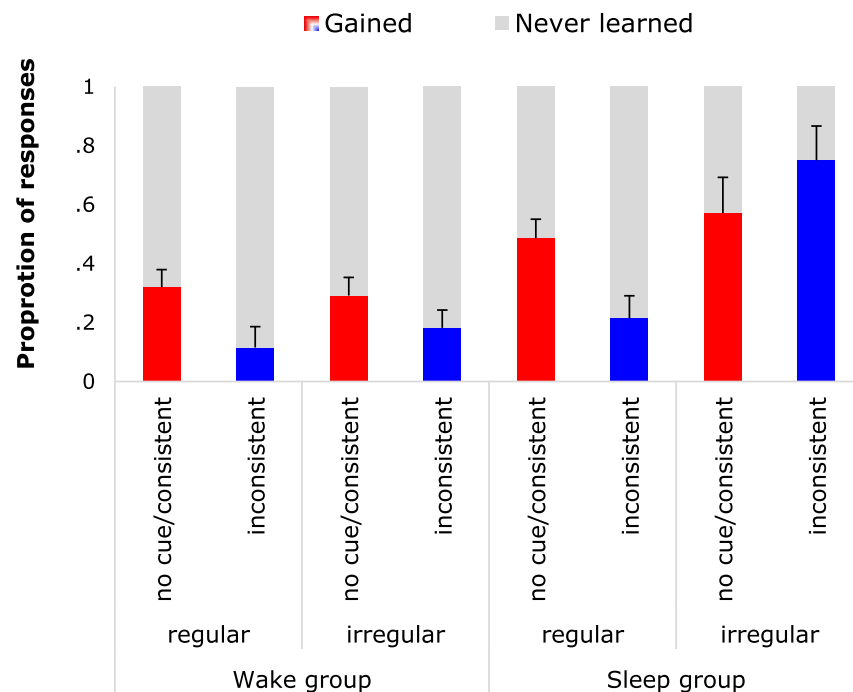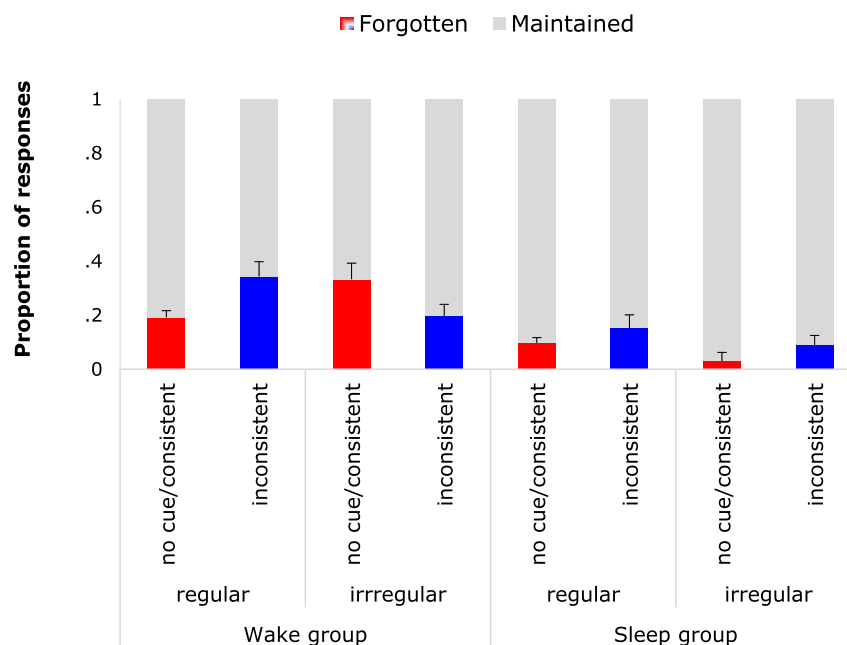

Fig. 11 — **Proportion of items that were forgotten across the delay for the two groups at cued recall. The proportions are calculated out of all items that were recalled correctly at the immediate session, and were then either forgotten (red and blue) or maintained (grey) at the delayed session.**
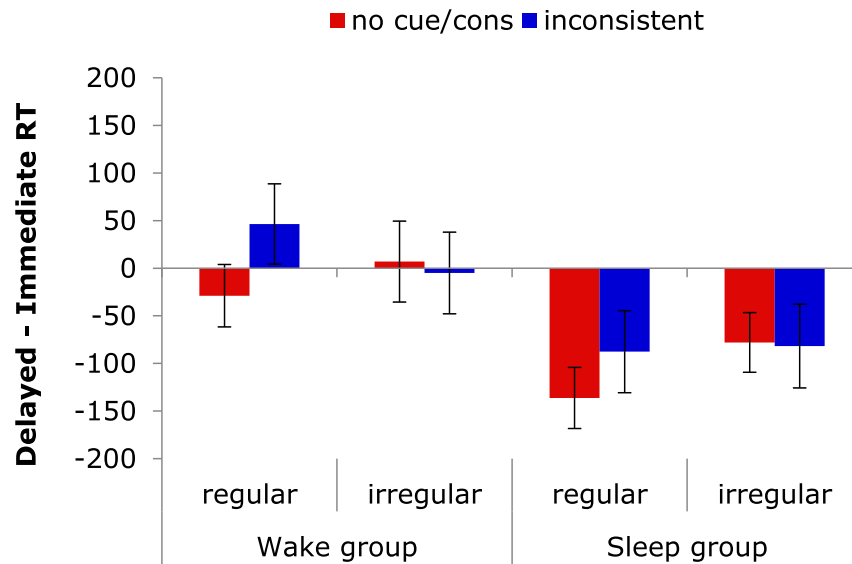
Fig. 12 — Recognition times on the 2AFC task illustrated using difference scores (delayed − immediate test) with trained plurals for the two groups.

(consistency: ß = .13, SE = .02, t = 5.22, p < .001), and more strongly so for the wake group (group x consistency: ß = −.06, SE = .03, t = −.98, p = .035). There were no other significant effects (Appendix 2, Table 6 and Fig. 2).

3.2.4. Relationship between trained items and generalization

To assess more directly the contribution of the memory of the trained items to generalization patterns in Experiment 2 we used a stepwise multiple regression. We focused on the generalization responses with the novel items with the ambiguous cue, as this was the condition where we saw balanced competition soon after learning in Experiment 1 and a clear change in performance across different delays, and in particular an increase in irregularizations. Hence we used the proportion of irregular inconsistent responses as the outcome measure. Our aim was to examine the extent to which both the initial learning and the consolidated knowledge of regular and irregular forms as measured by cued recall contributed to generalization performance. We were particularly interested in assessing the extent to which the strengthening of the memory for the trained irregular forms and weakening of the regular forms over the delay may have contributed to the pattern of generalization responses.

In the first two steps of the analysis, we assessed the contribution of initial learning of irregular inconsistent and then regular consistent forms. The model containing both predictors (model 2) provided an improvement over the model with only the initial learning of irregular inconsistent forms as the predictor (model 1) [F (1) = 5.74, p = .02; see Table 3 for model parameters]. In the next two steps, we added measures of consolidated knowledge (cued recall accuracy for the trained items at the delayed session) for irregular inconsistent (model 3) and then regular consistent forms (model 4). Again, each model provided an improvement over the models containing fewer predictors [model 3 vs model 2: F (1) = 10.32, p = .002; model 4 vs model 3: F (1) = 5.74, p = .02; Table 3]. The

addition of group (wake, sleep) as a predictor (model 5), did not provide any further improvements [model 5 vs model 4: F (1) = .31, p = .58].

These findings demonstrate that generalization performance was crucially influenced by the consolidated knowledge of the trained items. In particular, the stronger memory for the irregular inconsistent forms increased the likelihood of producing irregular inconsistent responses at generalization, and conversely the stronger memory for the regular inconsistent forms decreased the likelihood of producing irregularizations.

Table 3 — Model parameters with the proportion of irregular inconsistent generalizations for the ambiguous cue as the outcome measure.

predictors	beta	SE	t	p
model 1				
Initial learning: irregular inconsistent	.29	.12	2.32	.024
model 2				
Initial learning: irregular inconsistent	.39	.12	3.21	.002
Initial learning: regular inconsistent forms	−.38	.13	−2.85	.006
model 3				
Initial learning: irregular inconsistent	.20	.13	1.53	.131
Initial learning: regular inconsistent	−.40	.12	−3.21	.002
Consolidated: irregular inconsistent	.40	.12	3.21	.002
model 4				
Initial learning: irregular inconsistent	.24	.12	1.94	.058
Initial learning: regular inconsistent	−.15	.16	−.93	.355
Consolidated: irregular inconsistent	.49	.13	3.94	<.001
Consolidated: regular inconsistent	−.31	.13	−2.40	.020
model 5				
Initial learning: irregular inconsistent	.22	.13	1.75	.085
Initial learning: regular inconsistent	−.17	.17	−1.04	.303
Consolidated: irregular inconsistent	.51	.13	3.90	<.001
Consolidated: regular inconsistent forms	−.29	.13	−2.17	.035
Group	−.04	.07	−.56	.579

3.2.5. Summary

The aim of Experiment 2 was to explore further the contribution of different consolidation processes to the long-term retention of a quasi-regular inflectional system. In Experiment 1 we found that a delay of 24 h between training and generalization testing increased the production of irregular responses at generalization. In Experiment 2, we decreased the delay to 12 h, and we specifically focused on the contribution of sleep- and wake-related maintenance, strengthening, and forgetting of the newly acquired inflectional forms to the pattern of generalization responses.

We replicated the performance at initial acquisition of the novel inflectional system found in Experiment 1, including overall equally good learning of regular and irregular affixes. When tested on generalization of the newly learned inflections, the two 12 h groups again showed dominance of regularizations for no cue stems, and irregularizations for stems with the irregular consistent phonological cue. For the novel ambiguous stems for which competition between regular and irregular pluralization is most intense, we found that both groups showed an intermediate level of performance between the immediate and 24 h testing points from Experiment 1. There was an increase in the prevalence of irregular inconsistent responses at the two 12-h delays relative to the immediate testing point in Experiment 1 but this increase was not as strong as for the 24-h delayed group.

In Experiment 2 we also examined memory for the trained plurals across the delay and whether this memory could explain the generalization performance for the ambiguous stems. There were three key findings here. First, there was an overall advantage for the sleep over the wake condition in overall performance on the trained plurals. Across the 12-h delay, the sleep group gained more initially incorrectly recalled items than the wake group, and the wake group forgot more initially correct items than the sleep group. Second, there was an overall greater improvement for irregular forms over regular forms, with evidence in the forgetting analysis that this benefit was stronger for the sleep group than the wake group (and a nonsignificant ($p = .061$) effect in the gains analysis). Third, the consolidated knowledge of trained items contributed significantly to the performance on the generalization test above and beyond the initial learning levels. In particular, stronger consolidated memories of the irregular inconsistent trained forms increased the likelihood of producing irregularizations at generalization, and conversely stronger consolidated memories of regular forms decreased the likelihood of producing irregularizations.

It is also important to point out that the changes in the memory of the trained items over the delay period cannot be attributed to floor or ceiling effects. In both the wake and the sleep group participants initially performed better with consistent than inconsistent items, and this was particularly pronounced for regular items (Fig. 5): participants correctly recalled on average 87% of the regular consistent versus 68% of the regular inconsistent items in the sleep group, and 82% versus 72% respectively in the wake group (Fig. 1a, Appendix 2). The corresponding performance for the irregulars was more even (sleep group: 76% *vs* 77% for the consistent *vs* inconsistent irregulars respectively; wake group: 74% *vs*

76%). As reported above, we found evidence for a greater change in performance over the 12 h delay for regular than irregular items in both groups. There was clearly an opportunity in both groups and all conditions to reveal forgetting over the delay, while arguably there was less room for improvement (gains) in the regular consistent condition relative to other conditions. Crucially, with the average accuracy of 68% for the sleep group, and 72% for the wake group, this was not the case for the regular inconsistent items and yet we found few gains in either condition in the sleep group. Thus overall we believe that the analyses of gains and forgetting reported above provide a fair test of the changes in memory for the trained items over the 12 h delay.

4. General discussion

In the current study our aim was to examine the contribution of complementary learning systems to the acquisition, retention and consolidation of new mappings in a quasi-regular domain. We specifically assessed the extent to which different degrees of systematicity and arbitrariness in the mappings would influence learning and retention over 24 h.

The new quasi-regular system mimicked key properties of the English past tense in that the majority of the new forms were regular, and a minority were irregular. Our hypothesis was that new irregular, exceptional forms should particularly benefit from a hippocampal sparse encoding scheme, whereas more systematic regular mappings should be less reliant on the hippocampus and more able to exploit neocortical learning mechanisms. On the basis of this hypothesis, we predicted stronger consolidation effects for irregular than for regular forms.

Over two experiments, we found that both regular and irregular forms of this new inflectional system mimicking pluralization were learned well. Using tests of generalization of the new knowledge to previously unseen items, we found that immediately after learning, both types of forms were used at approximately equal rates. Novel singulars with no particular similarities to trained items showed a strong tendency to be captured by the type-dominant regular plural process, whereas novel singulars that were similar only to high frequency unique irregular items tended to be pluralized as if they were irregulars. These patterns were robustly manifested, and observed at all four time points tested (immediately after learning, after a 12 h delay including sleep or wake, and after a 24 h delay), suggesting that there was stability over time in the generalization process for these items. All this points to the establishment of a new inflectional system with many properties of real language systems such as the use of a default regular inflection that applies to a wide range of novel forms (a global form of generalization; cf. Xeroxed; Prasada & Pinker, 1993), but with irregularizations in cases where novel forms have strong similarities to clusters of previously learned irregular forms (e.g., *spling-splang*; Albright & Hayes, 2003; Marchman, 1997).

The most challenging items in our battery of novel singulars were the items that were placed in a part of the phonological space that had inconsistent pluralization, similar to

trained regular *and* irregular items. When generalization was tested for these items soon after learning, we found a roughly 50:50 split of regular and irregular responses. However, over 12-h and 24-h delays with no additional exposure to the new system, the irregulars became increasingly dominant in the generalization responses. This observation is consistent with the notion that the irregulars benefit more from consolidation processes than regulars, as predicted on the basis of our hypothesis.

However, contrary to our expectations this increasing influence of irregular items was not specifically related to sleep. Rather, the effect was observed equally strongly in the two 12-h groups with or without sleep. In order to try to understand the role of maintenance, strengthening, and forgetting of the trained items in explaining this behavior, we examined the recall of the trained plurals before and after the 12-h intervals including sleep and wake. This technique has been the subject of recent debate (Dumay, 2018; Schreiner & Rasch, 2016), but the incorporation of the technique into a mixed effects analysis proved revealing for our data. Consistent with many previous studies in word learning, we found overall poorer retention of the new inflectional system over the delay in the wake than the sleep group (see McMurray, Kapnoula, & Gaskell, 2017, for a review). Two general processes may have contributed to this outcome: we found stronger gains in recall for previously sub-threshold memories of the trained plurals in the sleep group, while at the same time there was greater overall forgetting of previously recalled trained plurals in the wake group. The overall better retention in the sleep group resulted from a combination of less forgetting (i.e., more robust maintenance) and also more gains in performance (perhaps due to sleep reactivating and strengthening initially weak memories).

Importantly, we also found that trained regular forms were less well retained over the 12-h interval than irregular forms. This effect was found regardless of whether the interval contained sleep or not, although we saw evidence of less forgetting for irregulars for the sleep but not for the wake group (indeed in both the gains and overall cued recall analysis the interaction with group was marginally significant; see Appendix 2, Tables 3 and 4). Putting these two effects together, we can see that the differential changes in memory for the trained plurals contribute to the shifts in generalization found for novel items. Irregular items were more likely to be strengthened (particularly in sleep) compared with regular items, whereas regular items were more likely to be forgotten (particularly in wake) than irregular items. In each case, the outcome was the same: a greater influence of the irregular trained materials in generalization to novel forms. Further evidence linking generalization performance to the retention of the training materials comes from a regression analysis, which showed that participants who retained strong memories of the irregular inconsistent plurals were more likely to produce irregular plurals of the ambiguous novel singulars, whereas participants who retained strong memories of the regulars were more likely to produce regular plurals. These effects were found even after controlling for post-training performance, suggesting that changes over the retention interval for trained items are key for explaining generalization performance.

Together, these findings highlight the importance of long-term memory processes in understanding the acquisition, retention, and use of new linguistic structures. We found that the knowledge of the quasi-regular inflectional system was not crystallized at the endpoint of the learning phase, with important changes taking place over the subsequent 24 h. It is reasonable to assume that these changes would continue to be influential over an even longer term as ongoing maintenance, strengthening, and forgetting processes progress over weeks or months.

It is worth noting that these changes were not observed for all the generalization materials that we tested. As mentioned, two properties (default regularization and generalization to consistent irregulars) were observed at all the time points tested. These are cases where the properties of the system did not conflict, and so any changes in memory for the trained materials did not have an observable effect. It was only when we focused on inconsistent forms for which there were strong competing tendencies to regularize and irregularize that we saw substantial changes in the generalization to new materials over time.

In many ways, the results conform with the hypotheses derived from the application of a complementary systems approach to language discussed in the introduction. The results can be explained in terms of a division of labor at encoding, with a greater reliance on the hippocampal system during the encoding of irregular forms, compared with a greater reliance on neocortical encoding of regular forms. Assuming that offline consolidation effects depend on the extent of hippocampal recruitment during learning, then this means that irregulars should be retained or strengthened to a greater extent than regulars over time. This was observed, both in terms of the retention of the trained items and the generalization to new materials.

That said, we also expected stronger maintenance and/or strengthening of hippocampally based memories across intervals including sleep compared with wake due to hippocampal replay (e.g., Skaggs & McNaughton, 1996; Staresina et al., 2015). We observed clear sleep benefits over wake in terms of the retention of the trained materials, including more robust maintenance of irregulars, but that did not translate into differential effects on generalization. Instead both 12 h including sleep and 12 h awake led to near identical shifts in the generalization pattern compared with the immediate test (12% for sleep and 13% for wake; see Fig. 8). Speculatively, the reason for the change in balance between regulars and irregulars across 12 h awake may relate to processes of forgetting impacting more on the regular, and particularly regular inconsistent forms, than the irregular forms (e.g., Sadeh et al., 2014; Sweegers & Talamini, 2014; Werchan & Gomez, 2014). Our finding of changes in generalization performance across both sleep and wake is part of a small but growing body of evidence that necessitates a better understanding of the offline processes that promote generalizable representations. Intriguingly, the effect of 24 h consisting of 12 h including sleep followed by 12 h awake on generalization (27%) was close to the summed effects of sleep and wake independently (12% + 13%). This could suggest that the wake and sleep effects are simply combined additively across a 24 h period, although we would be

in a better position to make that argument if we had also tested a delay of 24 h that included 12 h awake *prior to* sleep.

Although the current results demonstrate a differential effect of regularity over time that is broadly consistent with our predictions, it is worth considering an alternative explanation of the data. In order to set up an inflectional system that included a type-dominant regular alongside pockets of irregularity, we needed to ensure that the irregulars were dominant in terms of token frequency (Bybee & Slobin, 1982; Bybee, 1995b). Therefore, an alternative explanation of the current findings is that the same neural circuits were involved to the same extent in encoding all items (perhaps hippocampally mediated in all cases), but that over the offline period the processes of maintenance and strengthening favored the more robust memories (the high frequency plurals) over the less robust ones (the low frequency plurals), whereas forgetting favored the opposite. The literature here is mixed, but there is certainly evidence consistent with offline prioritization of certain types of memory during sleep (e.g., Oudiette, Antony, Creery, & Paller, 2013; Wilhelm et al., 2011; van Rijn, Lucignoli, Izura, & Blagrove, 2017). Therefore we cannot rule out the possibility on the basis of the current evidence that the factor governing the relative strengthening and weakening of memories over time is frequency rather than regularity. Nonetheless it is worth pointing out that recent evidence has suggested that weakly learned information might be prioritized during offline hippocampal replay, which would predict the opposite pattern of change from the one found (Schapiro, McDevitt, Rogers, Mednick, & Norman, 2017). Furthermore, in the current study there was no evidence that regularity influenced the strength of initial encoding: although there were some differences in performance on the trained plurals at the end of training, these were relatively small, and a main effect of regularity was not observed in either experiment. Finally, forgetting was strongest in both a set of high frequency (irregular consistent) and a set of low frequency (regular inconsistent) forms. Thus, while item frequency may play a role in initial encoding and long-term retention it does not easily account for the current findings.

4.1. Implications for models of language learning and use

Our findings have several implications for models of language learning and use, and in particular for debates about the processing of grammatical aspects of language (e.g., McClelland & Patterson, 2002; Pinker & Ullman, 2002; Seidenberg & Plaut, 2014). We have successfully mimicked the learning and generalization of an inflectional system in natural languages, in that participants learned both the default generalization pattern (global generalization), and 'islands of reliability' with predictable phonological cues (local generalizations; Albright & Hayes, 2003; Plunkett & Marchman, 1991, 1993).

The finding that generalization behavior was influenced by the memory of the trained regular *and* irregular forms is consistent with single-mechanism models of inflectional processing suggesting that both regular and irregular forms are processed within a single system encoding statistical regularities in the form-to-meaning mapping (e.g., Joanisse & Seidenberg, 1999; McClelland & Patterson, 2002; Seidenberg & Plaut, 2014). Moreover, the evidence that the memory for both

regular and irregular trained forms was influenced by domain-general memory consolidation processes lends further support to domain-general accounts of language learning and use (e.g., McClelland, 2015; Seidenberg, 1997).

Our findings are also relevant for the research that examines the type of information that contributes to grammatical generalizations (e.g., Endress & Hauser, 2011; Wonnacott, 2011; Wonnacott, Brown, & Nation, 2017; Wonnacott, Newport, & Tanenhaus, 2008). In the current study, generalization patterns immediately after learning have shown sensitivity to both regular, type-frequency based information in the input, and phonologically constrained cues (consistency in the mapping between the phonological cue in the stem and the affix). Crucially, offline memory consolidation processes have influenced representations of the learned input to increase the influence of irregular, token-frequency based generalizations over time. Thus future studies on the types of linguistic information influencing generalizations will need to take into account memory consolidation processes and how they shape grammatical generalizations.

Our experimental paradigm most clearly mimics morphological learning in a second language, in that participants were learning new words for existing concepts. Indeed, more recent dual-mechanism models suggest that morphological learning in a second language is better described as memory-based learning of all new forms, rather than only irregulars (e.g., Ullman, 2001). Our findings are also relevant for grammar learning in the first language in that, as shown by a number of recent studies, grammatical knowledge of the first language is malleable both in the short and the long term throughout the life-span (Fine, Jaeger, Farmer, & Qian, 2013; Kaschak, Kutta, & Schatschneider, 2011; Luka & Choi, 2012; Ryskin, Qi, Duff, & Brown-Schmidt, 2017; Wells, Christiansen, Race, Acheson, & MacDonald, 2009). The consolidation processes we described in the current study are thus likely to play a role across a range of phenomena in both first and second language acquisition.

Finally, our findings are relevant for understanding the learning and representation in quasi-regular domains in cognition more broadly, and in particular for computational models implementing domain-general mechanisms of learning and representation in quasi-regular systems (e.g., Armstrong et al., 2017; Harm & Seidenberg, 2004; Rogers & McClelland, 2004; Rogers & Patterson, 2007; Rogers et al., 2004). Our findings of consolidation related-changes suggest that these models, typically implementing the learning and representational mechanisms of the neocortical system may need to be augmented by explicitly implementing representational and learning mechanism of the hippocampal system and the interaction between the two. For example, Armstrong et al. (2017) have recently explored how the structure of representational space influences generalization when learning a similar quasi-regular system as in the current study but in the print-to-sound domain. Unlike the current study, their focus was on generalization performance at the consolidated state of the acquired knowledge (48 h after learning), and their behavioral findings were well matched with the predictions from their computational model implementing the distributed architecture of the neocortical system. Crucially, in order to avoid catastrophic interference when

learning new items [a known problem for distributed neocortical systems (c.f. McClelland et al., 1995)] the model implemented different error scaling rates for the existing versus the new items. For the purposes of their key research question examining the representational space in the neocortical system this may have been an appropriate simplification in the model, but the extent to which the same computational architecture would capture our findings of the important changes in the pre-consolidated knowledge remains to be tested in future studies.

4.2. Conclusions

Our study of the learning and retention of a new artificial morphological system over the course of 24 h has demonstrated the importance of considering the role of systematicity in the learning, consolidation and retention processes for verbal material. We found evidence that consolidation processes affected participants' ability to generate inflected forms of trained and novel stems, but these changes did not occur across the board. We found the strongest evidence of consolidation effects for uninflected items that had conflicting cues consistent with both regular (systematic) forms and irregular forms. In these cases increasing consolidation periods with or without sleep benefited irregulars over regulars. This result is broadly consistent with a complementary systems model in which encoding of nonsystematic irregular items relies on hippocampal pattern separation to a greater extent than for systematic items, and in which consolidation preferentially benefits the hippocampal memories. However, the finding that these changes in performance occur in sleep and wake equally calls for a better understanding of how sleep and wake combine to enhance the generalizability of knowledge.

Acknowledgments

Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.cortex.2018.07.015.

REFERENCES

Albright, A., & Hayes, B. (2003). Rules versus analogy in English past tenses: A computational/experimental study. *Cognition, 90*(2), 119–161. https://doi.org/10.1016/s0010-0277(03)00146-x.

Armstrong, B. C., Dumay, N., Kim, W., & Pitt, M. A. (2017). Generalization from newly learned words reveals structural properties of the human reading system. *Journal of Experimental Psychology General, 146*(2), 227–249. https://doi.org/10.1037/xge0000257.

Bakker, I., Takashima, A., van Hell, J. G., Janzen, G., & McQueen, J. M. (2014). Competition from unseen or unheard novel words: Lexical consolidation across modalities. *Journal of Memory and Language, 73*, 116–130. https://doi.org/10.1016/j.jml.2014.03.002.

Bakker, I., Takashima, A., van Hell, J. G., Janzen, G., & McQueen, J. M. (2015). Tracking lexical consolidation with ERPs: Lexical and semantic-priming effects on N400 and LPC responses to newly-learned words. *Neuropsychologia, 79*, 33–41. https://doi.org/10.1016/j.neuropsychologia.2015.10.020.

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). 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.

Batterink, L. J., Westerberg, C. E., & Paller, K. A. (2017). Vocabulary learning benefits from REM after slow-wave sleep. *Neurobiology of Learning and Memory, 144*, 102–113.

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

Bozic, M., & Marslen-Wilson, W. (2010). Neurocognitive contexts for morphological complexity: Dissociating inflection and derivation. *Language and Linguistics Compass, 4*(11), 1063–1073.

Brown, H., Weighall, A., Henderson, L. M., & Gaskell, M. G. (2012). Enhanced recognition and recall of new words in 7-and 12-year-olds following a period of offline consolidation. *Journal of Experimental Child Psychology, 112*(1), 56–72. https://doi.org/10.1016/j.jecp.2011.11.010.

Bybee, J. L. (1995). Diachronic and typological properties of morphology and their implications for representation. In L. B. Feldman (Ed.), *Morphological aspects of language processing*. Psychology Press.

Bybee, J. L. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes, 10*(5), 425–455. https://doi.org/10.1080/01690969508407111.

Bybee, J. L., & Slobin, D. I. (1982). Rules and schemas in the development and use of the English past tense. *Language, 58*(2), 265–289. https://doi.org/10.2307/414099.

Davis, M. H., & Gaskell, M. G. (2009). A complementary systems account of word learning: Neural and behavioural evidence. *Philosophical Transactions of the Royal Society of London B Biological Sciences, 364*(1536), 3773–3800. https://doi.org/10.1098/rstb.2009.0111.

Dewar, M., Alber, J., Butler, C., Cowan, N., & Della Sala, S. (2012). Brief wakeful resting boosts new memories over the long term. *Psychological Science, 23*(9), 955–960. https://doi.org/10.1177/0956797612441220.

Dewar, M., Alber, J., Cowan, N., & Della Sala, S. (2014). Boosting long-term memory via wakeful rest: Intentional rehearsal is not necessary, consolidation is sufficient. *PLoS One, 9*(10). https://doi.org/10.1371/Journal.Pone.0109542.

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

Dumay, N. (2016). Sleep not just protects memories against forgetting, it also makes them more accessible. *Cortex, 74*, 289–296. https://doi.org/10.1016/j.cortex.2015.06.007.

Dumay, N. (2018). Look more carefully: Even your data show sleep makes memories more accessible. A reply to Schreiner and Rasch. *Cortex, 101*, 288–293. https://doi.org/10.1016/j.cortex.2017.12.013.

Dumay, N., & Gaskell, M. G. (2007). Sleep-associated changes in the mental representation of spoken words. *Psychological Science, 18*(1), 35–39.

Endress, A. D., & Hauser, M. D. (2011). The influence of type and token frequency on the acquisition of affixation patterns: Implications for language processing. *Journal of Experimental Psychology Learning Memory and Cognition, 37*(1), 77—95. https://doi.org/10.1037/a0020210.

Fine, A. B., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PLoS One, 8*(10), e77661. https://doi.org/10.1371/journal.pone.0077661.

Forster, K. I., & Forster, J. C. (2003). DMDX: A windows display program with millisecond accuracy. *Behavior Research Methods Instruments and Computers, 35*(1), 116—124.

Friedrich, M., & Friederici, A. D. (2011). Word learning in 6-month-olds: Fast encoding—weak retention. *Journal of Cognitive Neuroscience, 23*(11), 3228—3240.

Friedrich, M., Wilhelm, I., Born, J., & Friederici, A. D. (2015). Generalization of word meanings during infant sleep. *Nature Communications, 6*.

Gais, S., Lucas, B., & Born, J. (2006). Sleep after learning aids memory recall. *Learning and Memory, 12*(1), 47—56. https://doi.org/10.1101/lm.132106.

Gaskell, M. G., & Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition, 89*(2), 105—132. https://doi.org/10.1016/s0010-0277(03)00070-2.

Glushko, R. J. (1979). The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology Human Perception and Performance, 5*(4), 674—691.

Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review, 106*(3), 491—528. https://doi.org/10.1037/0033-295x.106.3.491.

Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes. *Psychological Review, 111*(3), 662—720. https://doi.org/10.1037/0033-295X.111.3.662.

Hay, J. B., & Baayen, R. H. (2005). Shifting paradigms: Gradient structure in morphology. *Trends in Cognitive Sciences, 9*(7), 342—348. https://doi.org/10.1016/j.tics.2005.04.002.

Henderson, L. M., Weighall, A. R., Brown, H., & Gaskell, M. G. (2012). Consolidation of vocabulary is associated with sleep in children. *Developmenta Science, 15*(5), 674—687. https://doi.org/10.1111/j.1467-7687.2012.01172.x.

Himmer, L., Müller, E., Gais, S., & Schönauer, M. (2017). Sleep-mediated memory consolidation depends on the level of integration at encoding. *Neurobiology of Learning and Memory, 137*, 101—106.

Hoddes, E., Zarcone, V., Smythe, H., Phillips, R., & Dement, W. C. (1973). Quantification of sleepiness - new approach. *Psychophysiology, 10*(4), 431—436. https://doi.org/10.1111/J.1469-8986.1973.Tb00801.X.

Horst, J. S., & Samuelson, L. K. (2008). Fast mapping but poor retention by 24-month old infants. *Infancy, 13*(2), 319—335.

Jared, D. (1997). Spelling-sound consistency affects the naming of high-frequency words. *Journal of Memory and Language, 36*(4), 505—529.

Jared, D. (2002). Spelling-sound consistency and regularity effects in word naming. *Journal of Memory and Language, 46*(4), 723—750. https://doi.org/10.1006/jmla.2001.2827.

Jared, D., McRae, K., & Seidenberg, M. S. (1990). The basis of consistency effects in word naming. *Journal of Memory and Language, 29*(6), 687—715.

Joanisse, M. F., & Seidenberg, M. S. (1999). Impairments in verb morphology after brain injury: A connectionist model. *Proceedings of the National Academy of Sciences of the United States of America, 96*(13), 7592—7597. https://doi.org/10.1073/Pnas.96.13.7592.

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.

Kaschak, M. P., Kutta, T. J., & Schatschneider, C. (2011). Long-term cumulative structural priming persists for (at least) one week. *Memory and Cognition, 39*(3), 381—388. https://doi.org/10.3758/s13421-010-0042-3.

Kucker, S. C., & Samuelson, L. K. (2012). The first slow step: Differential effects of object and word-form familiarization on retention of fast-mapped words. *Infancy, 17*(3), 295—323. https://doi.org/10.1111/j.1532-7078.2011.00081.x.

Kumaran, D., Hassabis, D., & McClelland, J. L. (2016). What learning systems do intelligent agents need? Complementary learning systems theory updated. *Trends in Cognitive Sciences, 20*(7), 512—534. https://doi.org/10.1016/j.tics.2016.05.004.

Laudanna, A., Badecker, W., & Caramazza, A. (1992). Processing inflectional and derivational morphology. *Journal of Memory and Language, 31*(3), 333—348. https://doi.org/10.1016/0749-596x(92)90017-R.

Leminen, A., Kimppa, L., Leminen, M. M., Lehtonen, M., Makela, J. P., & Shtyrov, Y. (2016). Acquisition and consolidation of novel morphology in human neocortex: A neuromagnetic study. *Cortex, 83*, 1—16. https://doi.org/10.1016/j.cortex.2016.06.020.

Luka, B. J., & Choi, H. (2012). Dynamic grammar in adults: Incidental learning of natural syntactic structures extends over 48 h. *Journal of Memory and Language, 66*(2), 345—360. https://doi.org/10.1016/j.jml.2011.11.001.

Marchman, V. A. (1997). Children's productivity in the English past tense: The role of frequency, phonology, and neighborhood structure. *Cognitive Science, 21*(3), 283—303.

McClelland, J. L. (2013). Incorporating rapid neocortical learning of new schema-consistent information into complementary learning systems theory. *Journal of Experimental Psychology General, 142*(4), 1190—1210. https://doi.org/10.1037/a0033812.

McClelland, J. L. (2015). Capturing gradience, continuous change, and quasi-regularity in sound, word, phrase, and meaning. In B. MacWhinney, & W. O'Grady (Eds.), *The handbook of language emergence* (pp. 54—80). Hoboken, NJ: John Wiley & Sons.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review, 102*(3), 419—457.

McClelland, J. L., & Patterson, K. (2002). Rules or connections in past-tense inflections: What does the evidence rule out? *Trends in Cognitive Sciences, 6*(11), 465—472.

McClelland, J. L., & Rogers, T. T. (2003). The parallel distributed processing approach to semantic cognition. *Nature Reviews Neuroscience, 4*(4), 310—322. https://doi.org/10.1038/nrn1076.

McMurray, B., Kapnoula, E. C., & Gaskell, M. G. (2017). Learning and integration of new word-forms: Consolidation, pruning and the emergence of automaticity. In M. G. Gareth, & J. Mirković (Eds.), *Speech perception and spoken word recognition*. Psychology Press.

Mirković, J., & Gaskell, M. G. (2016). Does sleep improve your grammar? Preferential consolidation of arbitrary components of new linguistic knowledge. *PLoS One, 11*(4), e0152489. https://doi.org/10.1371/journal.pone.0152489.

Mirković, J., Seidenberg, M. S., & Joanisse, M. F. (2011). Rules versus statistics: Insights from a highly inflected language. *Cognitive Science, 35*(4), 638—681. https://doi.org/10.1111/j.1551-6709.2011.01174.x.

O'Reilly, R. C., & Rudy, J. W. (2001). Conjunctive representations in learning and memory: Principles of cortical and hippocampal function. *Psychological Review, 108*(2), 311—345.

Oudiette, D., Antony, J. W., Creery, J. D., & Paller, K. A. (2013). The role of memory reactivation during wakefulness and sleep in determining which memories endure. *Journal of Neuroscience, 33*(15), 6672—U6758. https://doi.org/10.1523/JNEUROSCI.5497-12.2013.

Pinker, S., & Ullman, M. T. (2002). The past and future of the past tense. *Trends in Cognitive Sciences, 6*(11), 456—463.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 103*(1), 56—115.

Plunkett, K., & Marchman, V. (1991). U-shaped learning and frequency effects in a multi-layered perceptron: Implications for child language acquisition. *Cognition, 38*(1), 43—102.

Plunkett, K., & Marchman, V. (1993). From rote Learning to system building: Acquiring verb morphology in children and connectionist nets. *Cognition, 48*(1), 21—69. https://doi.org/10.1016/0010-0277(93)90057-3.

Prasada, S., & Pinker, S. (1993). Generalisation of regular and irregular morphological patterns. *Language and Cognitive Processes, 8*(1), 1—56.

Rastle, K., Harrington, J., & Coltheart, M. (2002). 358,534 nonwords: The ARC nonword database. *Quarterly Journal of Experimental Psychology A, 55*(4), 1339—1362. https://doi.org/10.1080/02724980244000099.

Reid, A. M. (2013). *The influence of prior knowledge on memory consolidation* (PhD Thesis). University of York.

van Rijn, E., Lucignoli, C., Izura, C., & Blagrove, M. T. (2017). Sleep-dependent memory consolidation is related to perceived value of learned material in a Welsh language learning task. *Journal of Sleep Research, 26*, 302—308.

Rogers, T. T., Lambon-Ralph, M. A., Garrard, P., Bozeat, S., McClelland, J. L., Hodges, J. R., et al. (2004). Structure and deterioration of semantic memory: A neuropsychological and computational investigation. *Psychological Review, 111*(1), 205—235.

Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach.* Cambridge, Mass: MIT Press.

Rogers, T. T., & Patterson, K. (2007). Object categorization: Reversals and explanations of the basic-level advantage. *Journal of Experimental Psychology General, 136*(3), 451—469.

Rossion, B., & Pourtois, G. (2004). Revisiting snodgrass and vanderwart's object pictorial set: The role of surface detail in basic-level object recognition. *Perception, 33*(2), 217—236. https://doi.org/10.1068/p5117.

Ryskin, R. A., Qi, Z. H., Duff, M. C., & Brown-Schmidt, S. (2017). Verb biases are shaped through lifelong learning. *Journal of Experimental Psychology Learning Memory and Cognition, 43*(5), 781—794. https://doi.org/10.1037/xlm0000341.

Sadeh, T., Ozubko, J. D., Winocur, G., & Moscovitch, M. (2014). How we forget may depend on how we remember. *Trends in Cognitive Sciences, 18*(1), 26—36. https://doi.org/10.1016/j.tics.2013.10.008.

Schapiro, A. C., McDevitt, E. A., Rogers, T. T., Mednick, S. C., & Norman, K. A. (2017a). Human hippocampal replay during rest prioritizes weakly-learned information and predicts memory performance. *bioRxiv.* https://doi.org/10.1101/173021.

Schapiro, A. C., Turk-Browne, N. B., Botvinick, M. M., & Norman, K. A. (2017b). Complementary learning systems within the hippocampus: A neural network modelling approach to reconciling episodic memory with statistical learning. *Philosophical Transactions of the Royal Society B, 372*(1711), 20160049.

Schreiner, T., & Rasch, B. (2015). Boosting vocabulary learning by verbal cueing during sleep. *Cerebral Cortex, 25*(11), 4169—4179. https://doi.org/10.1093/cercor/bhu139.

Schreiner, T., & Rasch, B. (2016). To gain or not to gain - the complex role of sleep for memory: Comment on Dumay (2016). *Cortex.* https://doi.org/10.1016/j.cortex.2016.06.011.

Schreiner, T., & Rasch, B. (2017). The beneficial role of memory reactivation for language learning during sleep: A review. *Brain and Language, 167*, 94—105. https://doi.org/10.1016/j.bandl.2016.02.005.

Seidenberg, M. S. (1992). Connectionism without tears. In S. Davis (Ed.), *Connectionism: Theory and practice* (pp. 84—137). New York, NY: Oxford University Press.

Seidenberg, M. S. (1997). Language acquisition and use: Learning and applying probabilistic constraints. *Science, 275*(5306), 1599—1603. https://doi.org/10.1126/science.275.5306.1599.

Seidenberg, M. S., & Gonnerman, L. M. (2000). Explaining derivational morphology as the convergence of codes. *Trends in Cognitive Sciences, 4*(9), 353—361.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96*(4), 523—556.

Seidenberg, M. S., & Plaut, D. C. (2014). Quasiregularity and its discontents: The legacy of the past tense debate. *Cognitive Science, 38*(6), 1190—1228. https://doi.org/10.1111/cogs.12147.

Skaggs, W. E., & McNaughton, B. L. (1996). Replay of neuronal firing sequences in rat hippocampus during sleep following spatial experience. *Science, 271*(5257), 1870—1873. https://doi.org/10.1126/Science.271.5257.1870.

Staresina, B. P., Bergmann, T. O., Bonnefond, M., van der Meij, R., Jensen, O., Deuker, L., et al. (2015). Hierarchical nesting of slow oscillations, spindles and ripples in the human hippocampus during sleep. *Nature Neuroscience, 18*(11), 1679—1686. https://doi.org/10.1038/nn.4119.

Sweegers, C. C., & Talamini, L. M. (2014). Generalization from episodic memories across time: A route for semantic knowledge acquisition. *Cortex, 59*, 49—61. https://doi.org/10.1016/j.cortex.2014.07.006.

Tamminen, J., Davis, M. H., Merkx, M., & Rastle, K. (2012). The role of memory consolidation in generalisation of new linguistic information. *Cognition, 125*(1), 107—112. https://doi.org/10.1016/j.cognition.2012.06.014.

Tamminen, J., Davis, M. H., & Rastle, K. (2015). From specific examples to general knowledge in language learning. *Cognitive Psychology, 79*, 1—39. https://doi.org/10.1016/j.cogpsych.2015.03.003.

Tamminen, J., & Gaskell, M. G. (2008). Newly learned spoken words show long-term lexical competition effects. *Quraterly Journal of Experimental Psychology, 61*(3), 361—371. https://doi.org/10.1080/17470210701634545.

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. *Journal of Neuroscience, 30*(43), 14356—14360. https://doi.org/10.1523/JNEUROSCI.3028-10.2010.

Taraban, R., & McClelland, J. L. (1987). Conspiracy effects in word pronunciation. *Journal of Memory and Language, 26*(6), 608—631.

Tse, D., Langston, R. F., Kakeyama, M., Bethus, I., Spooner, P. A., Wood, E. R., et al. (2007). Schemas and memory consolidation. *Science, 316*(5821), 76—82. https://doi.org/10.1126/science.1135935.

Ullman, M. T. (2001). The neural basis of lexicon and grammar in first and second language: The declarative/procedural model. *Bilingualism Language and Cognition, 4*(2), 105—122. https://doi.org/10.1017/S1366728901000220.

Vlach, H. A., & Sandhofer, C. M. (2012). Fast mapping across time: Memory processes support children's retention of learned words. *Frontiers in Psychology, 3*, 46. https://doi.org/10.3389/fpsyg.2012.00046.

Wells, J. B., Christiansen, M. H., Race, D. S., Acheson, D. J., & MacDonald, M. C. (2009). Experience and sentence processing: Statistical learning and relative clause comprehension. *Cognitive Psychology, 58*(2), 250–271. https://doi.org/10.1016/j.cogpsych.2008.08.002.

Werchan, D. M., & Gomez, R. L. (2014). Wakefulness (not sleep) promotes generalization of word learning in 2.5-year-old children. *Child Development, 85*(2), 429–436. https://doi.org/10.1111/cdev.12149.

Wilhelm, I., Diekelmann, S., Molzow, I., Ayoub, A., Molle, M., & Born, J. (2011). Sleep selectively enhances memory expected to be of future relevance. *Journal of Neuroscience, 31*(5), 1563–1569. https://doi.org/10.1523/JNEUROSCI.3575-10.2011.

Williams, S. E., & Horst, J. S. (2014). Goodnight book: Sleep consolidation improves word learning via storybooks. *Frontiers in Psychology, 5*, 184. https://doi.org/10.3389/fpsyg.2014.00184.

Wojcik, E. H. (2013). Remembering new words: Integrating early memory development into word learning. *Frontiers in Psychology, 4.* https://doi.org/10.3389/Fpsyg.2013.00151.

Wonnacott, E. (2011). Balancing generalization and lexical conservatism: An artificial language study with child learners. *Journal of Memory and Language, 65*(1), 1–14. https://doi.org/10.1016/j.jml.2011.03.001.

Wonnacott, E., Brown, H., & Nation, K. (2017). Skewing the evidence: The effect of input structure on child and adult learning of lexically based patterns in an artificial language. *Journal of Memory and Language, 95*, 36–48. https://doi.org/10.1016/j.jml.2017.01.005.

Wonnacott, E., Newport, E. L., & Tanenhaus, M. K. (2008). Acquiring and processing verb argument structure: Distributional learning in a miniature language. *Cognitive Psychology, 56*(3), 165–209. https://doi.org/10.1016/j.cogpsych.2007.04.002.

Woollams, A. M., Joanisse, M., & Patterson, K. (2009). Past-tense generation from form versus meaning: Behavioural data and simulation evidence. *Journal of Memory and Language, 61*(1), 55–76.