



UNIVERSITY OF LEEDS

This is a repository copy of *Cross-modal working memory binding and L1-L2 word learning*.

White Rose Research Online URL for this paper:  
<http://eprints.whiterose.ac.uk/123394/>

Version: Accepted Version

---

**Article:**

Wang, S, Allen, RJ [orcid.org/0000-0002-1887-3016](https://orcid.org/0000-0002-1887-3016), Fang, S-Y et al. (1 more author) (2017) Cross-modal working memory binding and L1-L2 word learning. *Memory & Cognition*, 45 (8). pp. 1371-1383. ISSN 0090-502X

<https://doi.org/10.3758/s13421-017-0731-2>

---

**Reuse**

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



[eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk)  
<https://eprints.whiterose.ac.uk/>

## Cross-Modal Working Memory Binding and L1-L2 Word Learning

Shinmin Wang<sup>a</sup>, Richard J. Allen<sup>b</sup>, Shin-Yi Fang<sup>c</sup>, Ping Li<sup>c</sup><sup>a</sup> National Taiwan Normal University<sup>b</sup> University of Leeds<sup>c</sup> Pennsylvania State University

## Author Note:

Correspondence concerning this article should be addressed to Shinmin Wang,  
Department of Human Development and Family Studies, National Taiwan Normal  
University, 162, Sec.1, Heping E. Rd., Taipei 106, Taiwan. TEL: +886 (2)77341450. Email:  
s.wang@ntnu.edu.tw; Ping Li, Department of Psychology, Pennsylvania State University,  
University Park, PA 16802, USA. TEL: 814-863-3921. Email: pul8@psu.edu; Richard Allen,  
School of Psychology, University of Leeds, Leeds, LS2 9JT, UK. TEL: +44 (0)1133432667.  
Email: r.allen@leeds.ac.uk.

Shin-Yi Fang is now with the Sun Yat-Sen University, Guangzhou, China.

## CROSS-MODAL WORKING MEMORY AND WORD LEARNING

## Abstract

The ability to create temporary binding representations of information from different sources in working memory has recently been found to relate to the development of monolingual word recognition in children. The current study explored this possible relationship in an adult word learning context. We assessed whether the relationship between cross-modal working memory binding and lexical development would be observed in the learning of associations between unfamiliar spoken words and their semantic referents, and whether it would vary across experimental conditions in first and second language word learning. A group of English monolinguals were recruited to learn 24 spoken disyllable Mandarin Chinese words in association with either familiar or novel objects as semantic referents. They also took a working memory task in which their ability to temporarily bind auditory-verbal and visual information was measured. Participants' performance on this task was uniquely linked to their learning and retention of words for both novel objects and for familiar objects. This suggests that, at least for spoken language, cross-modal working memory binding might play a similar role in second language-like (i.e., learning new words for familiar objects) and in more native-like situations (i.e., learning new words for novel objects). Our findings provide new evidence for the role of cross-modal working memory binding in L1 word learning, and further indicate that early stages of picture-based word learning in L2 might rely on similar cognitive processes as in L1.

Key words: working memory, episodic buffer, cross-modal binding, word learning, second language

## CROSS-MODAL WORKING MEMORY AND WORD LEARNING

### Cross-Modal Working Memory Binding and L1-L2 Word Learning

Component processes central to the early stages of word learning include the mapping of pronunciations (phonological labels) to their printed words (orthographic forms) and semantic referents (e.g., visual objects for concrete nouns). This may initially involve the processing and temporary storage of phonological and visuospatial information both individually and in combination, within working memory. Previous work has already examined in detail the relationship between word learning and temporary storage and processing of information within isolated domains (i.e., visuospatial or verbal) (e.g., Baddeley, Gathercole, & Papagno, 1998; Gathercole, Willis, Emslie, & Baddeley, 1992; Majerus, Poncelet, Greffe, & van der Linden, 2006; Wang & Gathercole, 2013). However, few studies have investigated how the integration within working memory of information drawn across modalities might relate to word learning. A recent study based on written language attempted to address this issue, finding that children's capacity to temporarily bind novel auditory-verbal and visual information in working memory is linked to the development of their ability to form long-term orthography-phonology association in their native language (Wang, Allen, Lee, & Hsieh, 2015). What is as yet unknown is whether the mapping of spoken words to their semantic referents (e.g., objects for concrete nouns) during initial stages of word learning is also reliant on the ability to temporarily bind information across modalities in working memory in first (L1) language learning contexts; and if so,

## CROSS-MODAL WORKING MEMORY AND WORD LEARNING

whether it is also true in second (L2) language learning conditions. The aim of the current study is hence to extend the previous finding based on L1 written language to the domain of spoken language in an experimental simulation of L1 versus L2 new word learning in a young adult sample.

Working memory refers to a system that provides temporary maintenance of information necessary to support complex cognitive processing. A number of theoretical approaches have been developed in an attempt to capture working memory structure and function, and how it might relate to wider cognitive processes (e.g., Barrouillet, Bernardin, & Camos, 2004; Cowan, 1999; Unsworth & Engle, 2007). One of the most influential approaches is that advanced by Baddeley and colleagues (Baddeley, 1986, 1996; Baddeley, 2012; Baddeley & Hitch, 1974), with a multi-component model of working memory that includes separable phonological and visuospatial short-term stores and a central executive control process, along with the more recent addition of the episodic buffer, a modality-general store capable of integrating information drawn from different sources in the environment and from long-term memory (Baddeley, 2000; Baddeley, Allen, & Hitch, 2011). A common assumption across different approaches to working memory is that capacity is assumed to be extremely limited, with recently suggested estimates of around 3-5 items or chunks of information (e.g., Cowan, 2010). Basic temporary storage of verbal or visuospatial information is typically measured by simple span tasks that primarily require information retention, whereas complex span tasks

## CROSS-MODAL WORKING MEMORY AND WORD LEARNING

are designed to capture the simultaneous storage and processing of information. Using such measures, close links between working memory capacity and word acquisition have been identified (e.g., Baddeley et al., 1998; Gathercole et al., 1992; Majerus et al., 2006; Wang & Gathercole, 2013).

This earlier work has typically focused on temporary storage and processing of information within isolated domains (i.e. visuospatial or verbal), with little exploration of how the integration and binding within working memory of information drawn from different modalities might relate to word acquisition. Such exploration is potentially useful, given that learning mapping among phonology (i.e., pronunciations), orthography (i.e., printed words), and semantics (e.g., visual objects for concrete nouns) is an important process of early stage word learning (Seidenberg & McClelland, 1989). Indeed, research shows that the ability to learn arbitrary pairings between visual stimuli and phonological labels is a strong correlate of single word reading ability (e.g., Blomert, 2011; Hulme, Goetz, Gooch, Adams, & Snowling, 2007). However, it remains to be understood what might determine successful association formation between information from different modalities in long-term memory. Recent developments exploring temporary binding in working memory may provide insights into how temporary binding ability involved in initial stages of word learning might be linked to successful word acquisition (Jones, Branigan, Parra, & Logie, 2013; Wang et al., 2015). In the working memory literature, binding refers to the integration of individual features to form

## CROSS-MODAL WORKING MEMORY AND WORD LEARNING

a bound representation, typically following a single exposure (e.g., Allen, 2015; Brockmole & Franconeri, 2009).

One potentially useful theoretical framework to address this issue is that proposed by Baddeley and colleagues (Baddeley, 1986, 1996; Baddeley, 2012; Baddeley & Hitch, 1974), which specifies distinctions between visuospatial and phonological modalities, and the means to integrate such information within a modality-general episodic buffer (see also Barrouillet & Camos, 2014). This latter component is assumed to comprise a storage capacity based on a multidimensional code, which can be used to integrate information from specialized phonological and visuospatial subsystems, and to interface with long-term memory. This buffer may serve as a storage and modelling space that is informed by but separable from the specialized subsystems and long-term memory (Allen, Havelka, Falcon, Evans, & Darling, 2015; Langerock, Vergauwe, & Barrouillet, 2014), and may form an important stage in long-term episodic learning (Baddeley, 2003). The possible functioning of this proposed component has been intensively investigated by researchers using a range of tasks requiring memory judgments concerning combinations of features within domains (Allen, Baddeley, & Hitch, 2006; Allen, Baddeley, & Hitch, 2014; Brown & Brockmole, 2010; Hu, Hitch, Baddeley, Zhang, & Allen, 2014), between verbal and spatial domains (Elsley & Parmentier, 2009; Langerock et al., 2014; Morey, 2009), and across modalities (Allen, Hitch, & Baddeley, 2009). Outcomes to date indicate that ‘conjunctive’ or ‘intrinsic’ binding (e.g. of

## CROSS-MODAL WORKING MEMORY AND WORD LEARNING

shape and color within an object) may be relatively low-level and perceptual in nature, possibly accomplished by specialized visuospatial processing before being consciously retained within the episodic buffer. In contrast, ‘relational’ or ‘extrinsic’ binding of materials from different domains or modalities (as examined in the current study) may particularly require the episodic buffer for their formation and retention (see Allen, 2015 for a review; also, Ecker, Maybery, & Zimmer, 2013; Parra et al., 2013), as implied by Baddeley’s (2000) original proposal.

It would therefore be tempting to suggest that the episodic buffer concept may serve a useful purpose in understanding the possible links between temporary binding and word acquisition, through its proposed position at the interface of phonological processing, visuospatial processing, and long-term memory. Specifically, within this framework, verbal and visual information is assumed to be initially processed within specialized phonological and visuospatial subcomponents. However, memory for associations between information from these different subcomponents requires binding within the episodic buffer to form a unified representation. This proposed subcomponent within working memory may act as an intermediate store between initial maintenance and long-term storage, thus forming an important stage in the establishment of long-term representation of cross-modal associations. This perspective therefore suggests that the episodic buffer, or an equivalent modality-general aspect described in other theoretical approaches to working memory for example, Cowan’s



## CROSS-MODAL WORKING MEMORY AND WORD LEARNING

(1995, 2005) focus of attention, may play a role in the formation of short-term and longer-term phonological-semantic associations.

In order to start addressing this possible link, Wang et al. (2015) recently examined the performance of children, aged 8-9, on a task designed to measure their ability to bind information across auditory-verbal and visual modalities in working memory. This task assessed immediate memory for pairs of visually presented abstract shapes and auditorily presented nonwords, with children required to identify the original shape-nonword combinations they had encountered. Results showed that children's performance on this task significantly correlated with their single word recognition abilities even when chronological age, nonverbal ability, memory for individual features that constitute the binding task, and other reading-related factors were taken into account. These findings therefore suggest a unique link between children's working memory binding skills and their capacity to form long-term orthography-phonology associations.

The present study attempted to add to our understanding within this area by assessing the link between working memory binding ability and the mapping of phonological labels to their semantic referents (i.e., visual objects for concrete nouns) in an experimental simulation of L1 new word learning in young adults. In order to tap into the fundamental mental processes involved in L1 word learning, we used a task consisting of pairings of an auditorily presented novel word form (i.e. Mandarin Chinese) and a novel visual object depicting its

## CROSS-MODAL WORKING MEMORY AND WORD LEARNING

referent. The use of novel objects, that is, objects that participants have no experience with in daily life, could ensure the to-be-learned words would not be linked to already existing semantic referents in one's native language (Gupta, 2003). The use of Mandarin words that are phonologically-unfamiliar to our participants also help eliminate possible effects of long-term phonological knowledge on word learning results.

The second aim of the current study was to investigate whether the possible role of cross-modal working memory binding in formation of short-term and longer-term phonological-semantic association would similarly hold in second language learning. To investigate this issue, the present study also included a condition requiring learning of associations between novel words and familiar objects. Contrary to the novel word-novel object condition, learning novel phonological words for familiar objects would likely involve linking these novel labels to pre-existing word-semantic relations (Gupta, 2003). This resembles typical situations within second language learning, in which new labels (phonological word forms) are mapped onto existing concepts or percepts of objects and actions. Such distinctions between familiar versus novel objects have significant implications for the learning of a second language, as demonstrated in a study by Barcroft and Sunderman (2008) that compared the learning of associations between phonological forms and either familiar objects or novel objects. Their study suggested that novel words for familiar objects were more likely to be learned through L1 translation equivalents rather than through a direct

## CROSS-MODAL WORKING MEMORY AND WORD LEARNING

mapping of spoken words onto pictures that depict their referents (i.e., objects), in line with theoretical account of L1 and L2 word learning processes proposed by Kroll and Steward (1994) and Hernandez, Li, and MacWhinney (2005). According to these models, second language words are often mapped to labels of the first language (i.e., word-to-word associations) during early stages of learning, and direct word to concept links are only possible as the individual becomes more proficient in the second language. Contrary to this type of theoretical proposal, recent evidence has increasingly indicated that direct mappings between L2 words and their corresponding concepts are possible even for L2 learners who are still at an early point in their L2 learning (e.g., Poarch, van Hell, & Kroll, 2015). Based on results from this line of research, it might be expected that the capacity to bind auditory-verbal information to visual materials in working memory would also play a part in early stages of word learning for familiar objects (i.e., a second language-like situation). On the other hand, if L2 learners would rely exclusively on the lexical connections between L2 and L1 in their early stage of L2 new word learning, we would not expect to observe a link between cross-modal working memory binding performance and word learning outcomes in the learning condition of familiar objects.

In sum, the aims of this study are (1) to identify the role of cross-modal working memory binding in the mapping of spoken word forms to objects, and (2) to determine whether the relationship between cross-modal working memory binding and word learning is

## CROSS-MODAL WORKING MEMORY AND WORD LEARNING

modulated by the type of objects in the mapping process (familiar vs. novel objects). To achieve these aims, English monolinguals who had no knowledge of Mandarin Chinese were recruited to participate in a two-day training study. They learned 24 phonologically-unfamiliar novel words in Mandarin in association with either familiar or novel objects as semantic referents via novel word learning tasks (e.g., Hulme et al., 2007; H. Li, Shu, McBride-Chang, Liu, & Xue, 2009; Warmington & Hulme, 2012). Learning retention was subsequently assessed by delayed associative recognition tests administered one hour (T1) and then one day after initial learning (T2). They also completed a cross-modal working memory binding task (Wang et al., 2015), assessing immediate memory for auditory nonwords, for abstract shapes, and crucially, for the bindings between these features. Findings indicating a relationship between the ability to create and temporarily retain (for a few seconds) associations between auditory-verbal and visual stimuli in working memory and learning and retention (over 24 hours) of spoken words for novel objects would extend evidence for the role of working memory binding in long-term word learning based on written language (Wang et al., 2015) to the domain of spoken language. Furthermore, we wanted to assess how working memory binding performance might be related to learning of words for novel objects (a more native-like situation) versus for familiar objects (a second language-like situation), in order to identify whether cross-modal working memory binding might play a similar or differential role for new word learning in L1 and L2.

## Method

### Participants

Thus, A total of 71 English monolinguals were recruited from Pennsylvania State University, USA. None of the participants had prior experience with Mandarin Chinese or other tonal languages, as assessed by the Language History Questionnaire (P. Li, Zhang, Tsai, & Puls, 2014). Two participants responded by guessing randomly for more than half of the trials in the working memory binding memory task and their data were excluded from further analysis. Thus, data from 69 subjects (47 females; mean age 20.99, range 18-44,  $SD = 4.75$ ) were included in the present analyses. For the measures of associative recognition retention T1 (1-hour delay) and recognition retention T2 (1-day delay), two participants with a  $d'$  score of 0 or below was removed from that given measure at T1, as a value of 0 indicates inability to distinguish signal from noise and negative values of  $d'$  can also arise through response confusions (see Stanislaw & Todorov, 1999). At T2, two participants failed to come back for the recognition retention task, and one participant had a  $d'$  score of 0 or below in the novel-object condition. This left 66 and 67 participants in the novel-object condition and the familiar-object condition respectively. The participants took part in the experiment to gain credits for a course or to receive monetary compensation for their time. The study was approved by the Institutional Review Board of the Pennsylvania State University. Consent was obtained from each participant prior to the experiment.

## CROSS-MODAL WORKING MEMORY AND WORD LEARNING

### **Procedure**

The experiment was carried out on two consecutive days. On day 1, participants learned 24 disyllabic words in Mandarin in two conditions: in one condition 12 words were paired with familiar objects and in the other condition another 12 words were paired with novel objects. There was 3-minute break between conditions. After training, participants could take another 3-minute break. Participants then received a computerized working memory task that includes measures of memory for auditory nonwords, visual shapes, and crucially, for the binding between these features. This was followed by three further cognitive tasks. Results from these three additional tasks were not reported here because they were not focus of the current study. These three cognitive tasks took approximately 1 hour to finish. Afterwards, participants received an associative recognition test as the first delayed post-test (T1, 1-hour delay). On Day 2, participants came back to the lab and received the Raven's Progressive Matrices for nonverbal intelligence ability (Raven, Court, & Raven, 2006), and the other associative recognition test as the second delayed post-test (T2, 1-day delay). This associative recognition test was identical to that used at T1 except that the test pairs were presented in a different pre-determined random order.

### **Materials and Task**

**Word learning task.** In this task, 24 spoken disyllabic Mandarin words were either paired with familiar objects or novel objects, yielding a familiar-object condition and a novel-

## CROSS-MODAL WORKING MEMORY AND WORD LEARNING

object condition. The 24 spoken Mandarin words were adopted from a Mandarin word learning study that included a larger number of target words (Lan, Fang, Legault, & Li, 2015) (see Appendix A for the word list used). The spoken words were recorded by a female native Chinese speaker (see Lan et al. 2015 for details). The familiar and novel objects consisted of line-drawing pictures taken from Verma and Brysbaert (2015) (see Appendix B for a list of the pictures used). The familiar objects and novel objects were learned in separate blocks. Whether a spoken word was paired with familiar or novel objects was counterbalanced across participants, as was the presentation order of blocks.

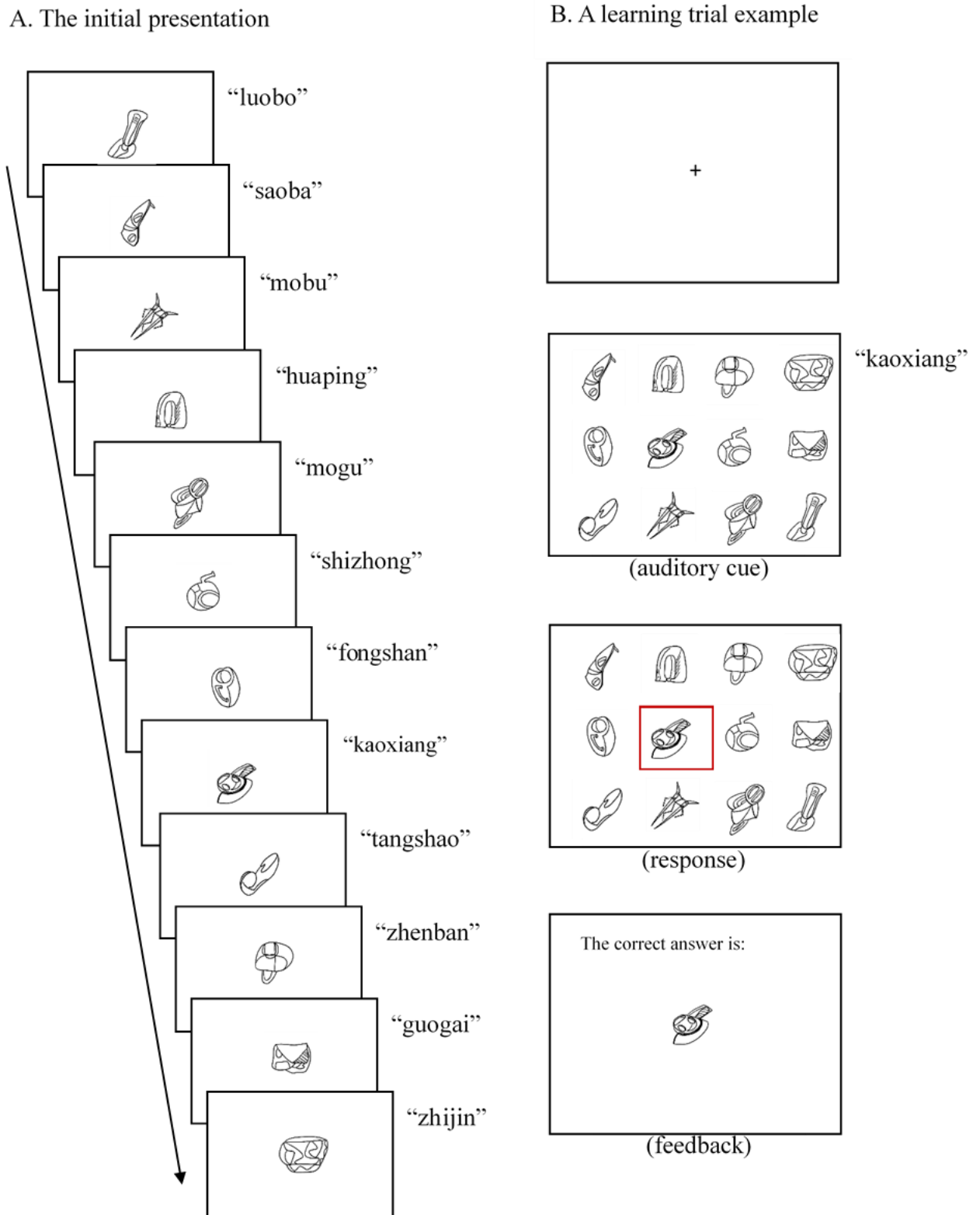
Each condition began with an initial presentation trial followed by five learning blocks, with each block containing 12 learning trials. In the initial presentation trial of each condition, 12 spoken word-object pairs were sequentially presented via a computer, one pair after another (see Figure 1A). Presentation time for each pair was 2 s. In the subsequent learning blocks, each learning trial (see Figure 1B for an example trial) started with a black fixation cross presented in the center of the screen. Once participants clicked on the fixation cross, one spoken Mandarin word was presented as a retrieval cue. Simultaneously, participants saw all 12 possible choices of picture items displayed evenly on the screen and were required to identify the target item by mouse clicking. Participants made a response within a self-determined time window (i.e., no time limit given). Once the response was made, the correct picture was shown on the next screen as feedback. Participants pressed the

## CROSS-MODAL WORKING MEMORY AND WORD LEARNING

spacebar to go to the next learning trial. Within each learning block, the 12 spoken Mandarin words were sampled randomly for each participant without replacement. The display locations of the 12 picture items were also randomized and changed across trials for each participant to prevent the use of location cues. After all 12 spoken Mandarin words were presented as retrieval cues in a learning block, another learning block began. All participants completed five learning blocks for the same 12 spoken word-object pairs. Proportion of correct responses was used as a dependent variable for each learning block, and mean proportion of correct responses across five learning blocks was used to index immediate learning outcomes for either learning condition. The entire training session on Day 1 lasted for approximately 30 minutes (including a 3-minute break between the two conditions).



## CROSS-MODAL WORKING MEMORY AND WORD LEARNING



**Associative recognition retention task.** Learning retention was assessed via an

## CROSS-MODAL WORKING MEMORY AND WORD LEARNING

associative recognition task using a procedure similar to that of Jackson and Schacter (2004).

There were 84 test pairs, consisting of 24 targets (intact pairs), 48 foils (24 rearranged pairs and 24 new pairs), and 12 additional trials (4 of the targets from either learning condition were repeated twice, and 4 other targets from either learning condition were repeated once).

The 24 targets were intact pairs that were presented unaltered from either training session.

The 48 foils consisted of 24 rearranged pairs constructed from separate study pairs and 24 entirely new pairs where both the constituent elements did not appear in the training sessions.

The inclusion of entirely new pairs would allow us to assess whether participants exhibited successful item memory. To create 24 entirely new pairs, 24 new spoken Mandarin words and 24 new picture items were included. The 24 new Spoken Mandarin words (see Appendix A for a list of words) were also adopted from Lan et al. (2015). The 24 new picture items (12 real and 12 novel) were drawn from the study of Verma and Brysbaert (2015) (see Appendix B for a list of pictures). Finally, the additional 12 target pairs were included to minimize the possibility that participants would base their responses on how they had responded to similar items that had been presented earlier in the test (Bayley, Wixted, Hopkins, & Squire, 2008; Holdstock et al., 2002). Hence, for example, if a participant had already seen pair A-B, he or she could not assume that any subsequent pairs that was composed of A or B would necessarily be a rearranged pair. Scores were based on responses to the first occurrence of the target pairs (Holdstock et al., 2002).

## CROSS-MODAL WORKING MEMORY AND WORD LEARNING

The 84 pairs were presented one at a time in a pre-determined random order. Participants heard the sound of a Chinese word and simultaneously saw an object picture shown in the center of the screen. They had to indicate their memory for each pair by pressing one of four keys corresponding to four response options that are “intact”, “rearranged”, “new” or “single”. Descriptions of the response option of “intact”, “rearranged” and “new” were shown above. The “single” option denotes the pairs that are composed of a word/object from the study and a new word/object. Note that although the test included a “single” response option, there were no test pairs of this type. “Single” response options were designed to avoid the possible strategy that participants may begin to make “rearranged” responses after recognizing only a single item of the pair, and thus contaminating “rearranged” responses (see Jackson & Schacter, 2004, for a similar design).

There was no time limit for responding. The same associative recognition task was used as delayed post-test at T1 and T2, with test pairs presented in a different and pre-determined random order that was used for all participants at the two time points. Dependent variables were calculated as  $d'$  scores based on the response to the first occurrence of 24 target pairs and on the response to the 24 rearranged pairs.<sup>1</sup>

**Cross-modal working memory binding task.** The computerized binding task designed by Wang et al. (2015) was adjusted to measure working memory for auditory-verbal and visual information binding (binding condition). In order to be able to separate participants'

## CROSS-MODAL WORKING MEMORY AND WORD LEARNING

memory capacity for individual features from their ability to form associations between features in working memory, two corresponding feature memory tasks were also administered to measure memory for constituent auditory-verbal materials (verbal condition) and visual materials (visual condition).

The stimuli consisted of a set of 8 English auditory nonwords (fren, bris, cral, tros, drup, srap, prin, grol) used by Jones, Branigan, Parra and Logie (2013) and a set of 8 abstract and non-nameable six-point shapes (numbers 3, 7, 13, 19, 20, 21, 29, 30) drawn randomly from the study of Vanderplas and Garvin (1959). The auditory nonwords were recorded by a male English speaker. The visual shapes were presented in black against a white background, with an approximate size of 1.6 cm<sup>2</sup> (see Figure 2 for examples). All stimuli were sampled randomly without replacement within each trial and used for all participants.

Each memory condition (auditory-verbal, visual, binding) was presented in a separate block. The tasks began with the presentation of the auditory-verbal or visual condition, counterbalanced across participants. To ensure that auditory-verbal materials and visual materials were equally familiar to participants in the binding condition, the binding condition was always administered as the final task (following a 3-minute break after the feature conditions). List length was set at 3 items in the binding condition and 6 items in the auditory-verbal and visual conditions, to equate the number of individual features to be remembered between binding and individual memory conditions (Wang et al., 2015). Each

## CROSS-MODAL WORKING MEMORY AND WORD LEARNING

condition contained 15 experimental trials, preceded by 4 practice trials of a 2-item sequence.

In the auditory-verbal condition (Figure 2A), at the study phase, a sequence of 6 auditory nonwords was presented via headphones. Each trial began with a black fixation cross presented at the upper centre of the screen for 500ms followed by a 250ms delay. Each to-be-remembered item was then presented for 1000ms, with inter-stimulus intervals of 250ms, with the screen remaining blank during presentation. A 1000ms delay followed offset of the final item in the sequence and was then followed by the test phase.

At the test phase, all 8 nonwords were displayed in their visual forms (fren, bris, cral, tros, drup, srup, prin, grol) as response options in the lower half of the screen, each surrounded by a grey square outline. The participants used the mouse to click the target items in any order. No serial order element was required as this was not an explicit part of the binding task. The grey square around the items turned green once selected and remained green till the end of the test phase as a reminder of which items had been selected. The next trial started automatically once all responses had been made or when the total response time exceeded 36s, giving 6s on average for each response. Display locations of the 8 response options at test were randomized and changed across trials to prevent the potential use of location cues. The dependent variable was proportion of correct responses.

In the visual condition (Figure 2B), the procedure was identical to that employed in the auditory-verbal condition, except that the experimental stimuli were replaced by a set of 8

## CROSS-MODAL WORKING MEMORY AND WORD LEARNING

shapes (described above). At study, a sequence of 6 shapes was presented at the upper centre of the screen at study. At test, the 8 possible choices were presented in the lower half of the screen. The participants had to click to select the target items in any order. The dependent variable was proportion of correct responses.

In the binding condition (Figure 2C), at study, a sequence of 3 arbitrary pairs of auditory-verbal and visual stimuli was presented. The task procedure was identical to that employed in the auditory-verbal condition and visual condition, except that the presentation time for the constituent shape of each to-be-remembered pair was extended to 2000ms. The presentation time for each auditory nonword remained the same as that in the feature condition (i.e., 1000ms). This gave participants 2000ms to process each pair, and hence ensured equivalent feature processing for the individual feature and binding memory conditions. At test, either auditory nonwords or visual shapes were presented one at a time as retrieval cues. Simultaneously, participants saw all 8 possible choices of the other features that made up the pair in the study phase displayed at the lower half of the screen, and were required to identify the target item by mouse clicking. On the trials where visual shapes were presented as retrieval cues, auditory nonwords were displayed in their visual forms as response options.<sup>2</sup> The maximum response time for each cue was 6s. To prevent the role of serial order mechanism, cue items were randomly presented on each given trial. The dependent variable was proportion of correct responses. Additionally, guessing error rate was

## CROSS-MODAL WORKING MEMORY AND WORD LEARNING

examined as a proportion of the total number of features that did not appear in the presented sequence but were selected. This response type represents participants' tendency to produce random guesses when performing the binding task (two participants' data were removed from analysis due to random guesses; see Participants).

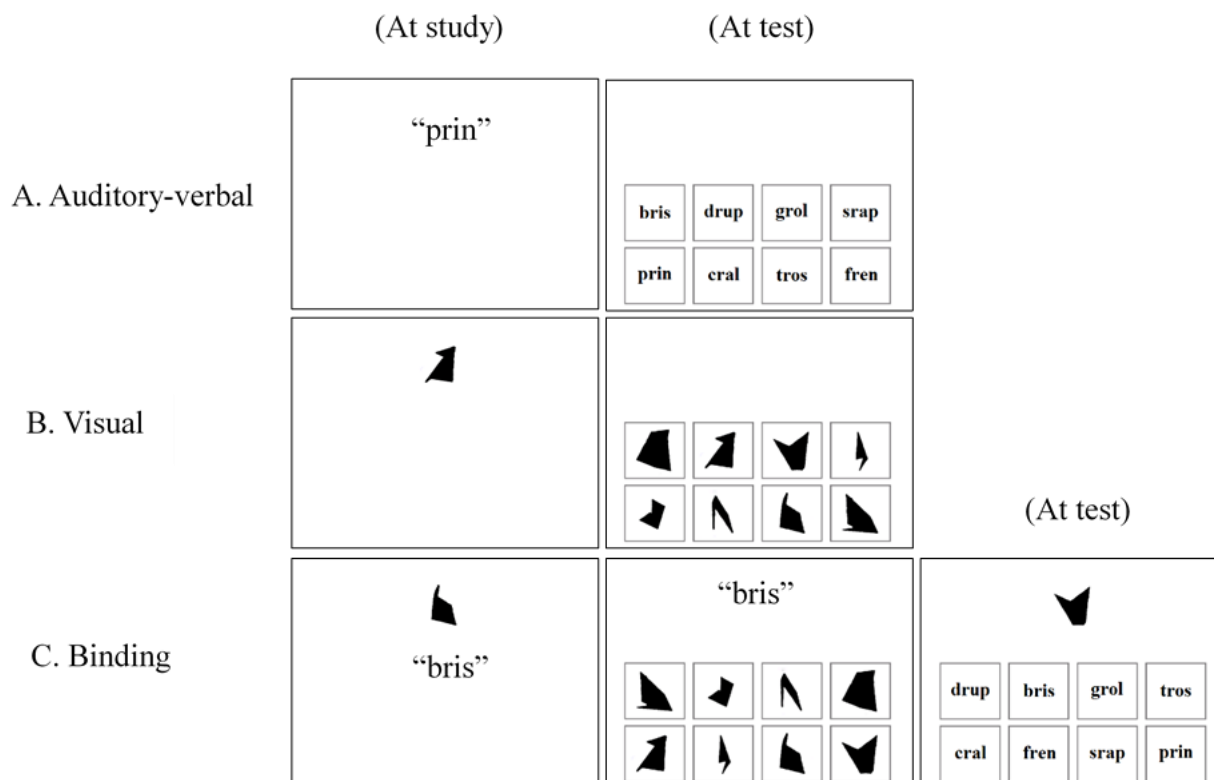


Figure 2. Cross-modal working memory binding task.

## Results

### Learning and Retention Performance

For the initial learning outcomes, mean proportion of correct responses for the five learning blocks were .21(SD=.14), .36(SD=.19), .46(SD=.22), .62(SD=.23), .67(SD=.25) in the novel-object condition

## CROSS-MODAL WORKING MEMORY AND WORD LEARNING

and .34(SD=.18), .47(SD=.24), .60(SD=.26), .69(SD=.22), .80(SD=.19) in the familiar-object condition. In each condition, participants' mean performance levels across learning blocks were averaged as dependent variables and used in the following analyses. The descriptive statistics for all principal measures are shown in Table 1.

Table 1  
Descriptive Data for Principle Measures

	Initial Learning Outcomes (proportion correct)		Associative Recognition Retention T1 (d')		Associative Recognition Retention T2 (d')		Working Memory Binding (proportion correct)			Nonverbal Ability (raw, Max.=60)
	Novel	Real	Novel	Real	Novel	Real	Binding	Auditory	Visual	
Mean	0.45	0.55	1.93	2.51	1.92	2.32	0.55	0.84	0.86	48.12
SD	0.17	0.19	0.91	0.78	0.86	0.88	0.15	0.05	0.06	6.52
Min.	0.15	0.13	0.21	0.43	0.21	0.43	0.17	0.72	0.68	26
Max.	0.82	0.95	3.46	3.46	3.46	3.46	0.87	0.93	0.97	59
N	69	69	67	67	66	67	69	69	69	69

Note. 1. Dependent variable of learning was proportion of correct responses averaged across 5 learning blocks.

2. T1 (1-hr after learning); T2 (1-day after learning).

The distribution of associative recognition responses is presented in Table 2. Across the 4 conditions (novel-object conditions across T1 and T2; familiar-object conditions across T1 and T2), participants correctly recognized .75 -.87 of all intact pairs. Very few rearranged pairs were mistaken for intact pairs (.10 -.16). These results demonstrated that participants' achieved relatively accurate associative memory representations. Recognition of individual items was also evidenced by the fact that most rearranged pairs were correctly identified (.74 -.80), while few new pairs were labelled 'rearranged' (1%-3%). In the absence of associative recognition for intact pairs, participants were more likely to make a 'rearranged' response (.11 -.19) than 'new' (.00-.01) or 'single' (.02 -.06) responses, suggesting these responses were also driven by successful item memory. These results, taken together, suggest that the participants demonstrated success in the learning and retention of the target pairs.



## CROSS-MODAL WORKING MEMORY AND WORD LEARNING

Discriminability scores ( $d'$ ) were respectively calculated for associative recognition retention tasks across the 4 conditions (novel-object conditions across T1 and T2; familiar-object conditions across T1 and T2) (Table 1) and used for the following correlational analyses. Rates of hits and false alarms were calculated based on participants' responses to the intact trials and rearranged trials. When the hit or false-alarm rate equaled 0 or 1, rates of 0 were converted to  $0.5/n$ , and rates of 1 to  $(n-0.5)/n$ , where  $n$  is the number of signal or noise trials (Macmillan & Kaplan, 1985).

Table 2  
Average Distribution Proportion of Associative Recognition Responses Sorted by Test Item Type

Time	Test response	Novel Object			Familiar Object		
		Test item category			Test item category		
		Rearranged (12 pairs)	Intact (12 pairs)	New (12 pairs)	Rearranged (12 pairs)	Intact (12 pairs)	New (12 pairs)
T1	"Rearranged"	0.75	0.18	0.02	0.80	0.11	0.02
	"Intact"	0.14	0.75	0.00	0.10	0.87	0.00
	"New"	0.01	0.01	0.82	0.00	0.00	0.78
	"Single"	0.10	0.06	0.17	0.10	0.02	0.19
T2	"Rearranged"	0.74	0.19	0.03	0.76	0.12	0.01
	"Intact"	0.16	0.76	0.00	0.14	0.85	0.00
	"New"	0.02	0.01	0.85	0.01	0.00	0.85
	"Single"	0.09	0.05	0.13	0.09	0.03	0.13

Note. T1 (1-hr after learning); T2 (1-day after learning).

### Association between Cross-modal Working Memory Binding and Learning

#### Performance

The simple correlation between measures is shown in Table 3. One-tailed tests were used in the following statistical significance testing given that we hypothesized a positive correlation between binding memory and word learning. The simple correlations show that the measure of cross-modal working memory binding was significantly correlated with the

## CROSS-MODAL WORKING MEMORY AND WORD LEARNING

learning and retention of word-object pairs in all conditions across different time points

( $r=.287\sim.449$ ).

Table 3  
Correlations between Measures

	1	2	3	4	5	6	7	8	9	10
1. Initial Learning Outcomes (novel)	1									
2. Initial Learning Outcomes (familiar)	.349**	1								
3. Associative Recognition Retention T1 (novel)	.746**	.434**	1							
4. Associative Recognition Retention T1 (familiar)	.345**	.632**	.427**	1						
5. Associative Recognition Retention T2 (novel)	.673**	.319**	.730**	.242*	1					
6. Associative Recognition Retention T2 (familiar)	.303**	.724**	.434**	.632**	.353**	1				
7. Binding Memory	.364**	.449**	.349**	.287**	.314**	.334**	1			
8. Auditory Memory	.198	.248*	.252*	.023	.352**	.241*	.255*	1		
9. Visual Memory	.241*	.400**	.235*	.297**	.224*	.346**	.375**	.278*	1	
10. Noverbal Ability	.183	.122	.058	.007	.099	-.004	.296**	.115	.266*	1

Note. T1 (1-hr after learning); T2 (1-day after learning); \*\*  $p < .01$  (1-tailed); \*  $p < .05$  (1-tailed).

A set of hierarchical regression analyses was then carried out with word learning outcomes across conditions and time points as dependent variables (Table 4). For each analysis, nonverbal ability, auditory-verbal memory and visual memory were entered at step 1, to ensure that any observed correlations between the working memory binding task and word learning were not simply due to effects of general cognitive resources and individual differences in memory capacities for individual features. Results revealed that working memory binding was a significant predictor of successful word learning in the novel-object conditions across the three time points (initial learning outcome; associative recognition retention at T1; associative recognition retention at T2), accounting for a unique 6.6%, 6.8% and 4.1% of variance in word learning performance, respectively. For the familiar object conditions, the working memory binding task was also a significant predictor, accounting for 9.5%, 5.1%, and 5.4% of variance in word learning performance at the respective time points.

RUNNING HEAD: CROSS-MODAL WORKING MEMORY AND WORD LEARNING

Table 4  
Hierarchical Regressions

Step	Independent variables	Dependent variables													
		Initial Learning Outcomes (Novel)							Initial Learning Outcomes (Familiar)						
		b (CI 95%)	SE B	$\beta$	p	Total R <sup>2</sup>	$\Delta R^2$	p	b (CI 95%)	SE B	$\beta$	p	Total R <sup>2</sup>	$\Delta R^2$	p
	Constant	-0.305 (-1.127, 0.516)	0.411		.230				-0.617 (-1.468, 0.235)	0.426		.076			
1	Nonverbal Ability	0.002 (-0.005, 0.008)	0.003	.063	.303	.034	.034	.066	-0.002 (-0.009, 0.005)	0.003	-.061	.298	.015	.015	.159
2	Auditory Memory	0.335 (-0.552, 1.222)	0.444	.092	.227	.091	.057	.070	0.388 (-0.531, 1.307)	0.460	.095	.201	.181	.166	.001
	Visual Memory	0.250 (-0.453, 0.952)	0.352	.091	.240				0.798 (0.070, 1.526)	0.364	.261	.016			
3	Binding Memory	0.331 (0.035, 0.627)	0.148	.288	.015	.156	.066	.015	0.444 (0.138, 0.750)	0.153	.345	.003	.276	.095	.003
	Associative Recognition Retention (Novel, T1)							Associative Recognition Retention (Familiar, T1)							
		b (CI 95%)	SE B	$\beta$	p	Total R <sup>2</sup>	$\Delta R^2$	p	b (CI 95%)	SE B	$\beta$	p	Total R <sup>2</sup>	$\Delta R^2$	p
	Constant	-2.262 (-6.592, 2.069)	2.166		.150				1.087 (-2.650, 4.824)	1.870		.282			
1	Nonverbal Ability	-0.010 (-0.044, 0.024)	0.017	-.072	.282	.003	.003	.320	-0.015 (-0.045, 0.015)	0.015	-.127	.157	.000	.000	.477
	Auditory Memory	3.020 (-1.645, 7.686)	2.334	.159	.100	.093	.089	.026	-1.621 (-5.660, 2.418)	2.020	-.099	.213	.097	.097	.020
	Visual Memory	1.382 (-2.400, 5.163)	1.892	.096	.234				3.234 (0.038, 6.430)	1.599	.265	.024			
2	Binding Memory	1.748 (0.190, 3.305)	0.779	.293	.014	.161	.068	.014	1.314 (-0.056, 2.684)	0.685	.251	.030	.148	.051	.030
	Associative Recognition Retention (Novel, T2)							Associative Recognition Retention (Familiar, T2)							
		b (CI 95%)	SE B	$\beta$	p	Total R <sup>2</sup>	$\Delta R^2$	p	b (CI 95%)	SE B	$\beta$	p	Total R <sup>2</sup>	$\Delta R^2$	p
	Constant	-3.743 (-7.814, 0.328)	2.036		.036				-2.575 (-6.640, 1.490)	2.033		.105			
1	Nonverbal Ability	-0.003 (-0.035, 0.029)	0.016	-.024	.423	.010	.010	.215	-0.023 (-0.054, 0.009)	0.016	-.170	.082	.000	.000	.486
	Auditory Memory	5.070 (0.701, 9.439)	2.185	.281	.012	.143	.134	.006	2.439 (-1.929, 6.807)	2.185	.132	.135	.156	.156	.003
	Visual Memory	1.013 (-2.431, 4.457)	1.722	.075	.280				3.646 (0.200, 7.092)	1.724	.265	.019			
2	Binding Memory	1.289 (-0.175, 2.752)	0.732	.227	.042	.185	.041	.042	1.503 (0.047, 2.959)	0.728	.259	.022	.211	.054	.022

Note. T1 (1-hr after learning); T2 (1-day after learning).

### Discussion

The goal of this study was to enhance our understanding of the relationship between the ability to temporarily bind information across modalities in working memory and novel word learning in L1 and L2. After controlling for general cognitive abilities and memory for individual features, the results showed that individuals' ability to bind auditory-verbal and visual information for immediate response in a working memory task was a unique predictor of learning and retention in long-term memory of phonologically-unfamiliar words for novel objects (a context resembling L1 word learning processes) and for familiar objects (a context resembling L2 word learning processes). The finding that cross-modal working memory binding performance uniquely predicted success in the mapping of phonological labels to novel objects thus extends previous findings based on written language (Wang et al., 2015) to the domain of spoken language in the L1 learning context. Moreover, the current results suggest that early stage L2 word learning is also reliant on cross-modal working memory binding to some extent.

In the case of learning phonological labels for novel objects, performance on the working memory binding task where participants were required to bind information across different modalities (and retain it for a few seconds) significantly predicted initial word-object pair learning outcomes, and importantly, retention performance measured approximately 1 hour and then 1 day after learning. Crucially, regression analyses indicate

## CROSS-MODAL WORKING MEMORY AND WORD LEARNING

that this link cannot be readily explained by general cognitive skills and memory for individual materials that constitute the binding task. That such a relationship can be observed between tasks examining retention over very different timescales, and using different stimuli and response measures, fits well with the idea that temporary feature-binding ability within working memory may form an important stage in long-term episodic learning (Baddeley, 2003), and is also consistent with Wang et al.'s (2015) finding that the capacity to temporarily integrate auditory-verbal and visual information in working memory is linked to long-term word acquisition in children. It is worth noting that the observed link between working memory binding performance and retention of the learned word-object pairs does not seem to simply arise from similarity between test formats, as the working memory binding task employed a reconstruction paradigm while the associative retention tasks used a single-probe procedure. These results are therefore in line with previous work indicating that learning novel words for objects that are previously unnamed, as is the case for early stages of L1 learning, would force learners to directly map the spoken words to their corresponding objects (Barcroft & Sunderman, 2008). As such, it would inevitably rely on the ability to bind information across verbal and visual modalities.

In addition, the current study demonstrated that cross-modal working memory binding performance was also associated with learning and retention of 'novel word-familiar object' pairings. This finding runs counter to the theoretical assumption that L2 word forms are

## CROSS-MODAL WORKING MEMORY AND WORD LEARNING

exclusively attached to the memory system by lexical links with L1 rather than through direct conceptual links during early stages of L2 learning (Hernandez, Li, & MacWhinney, 2005; Kroll & Stewart, 1994), as they are already named in learners' first language. The predictive power demonstrated by cross-modal binding in 'novel word-familiar object' word learning outcomes, on the contrary, may imply that the direct mapping of L2 words onto their semantic referents (i.e. concepts) are possible even for L2 learners who are still at early point in their L2 learning (e.g., Poarch, van Hell, & Kroll, 2015).

We would note that our participants learned 'L2' words via the picture-word paired-associate learning method (i.e., a picture-based method). The use of images may encourage a strategy of direct mapping between phonological and visual information to establish novel word-familiar object pairings, which could be different from other strategies such as L1 word-L2 word association. It is also worth noting that the findings from Poarch et al. (2015) were based on child L2 learners of English in the Netherlands whose word learning contexts were enriched by pictures and speaking/listening experiences, which is unlike the typical context encountered by adult L2 learners. In future studies, it may be productive to examine whether different types of learning method (e.g., picture-based vs. word-based method) and learning context (e.g., associative learning vs. immersed learning context) would mediate the role of cross-modal working memory binding in early stages of L2 word learning. It may also be worthwhile to assess whether different forms of working memory binding (e.g., word-

## CROSS-MODAL WORKING MEMORY AND WORD LEARNING

nonword binding as well as cross-modal binding) differentially contribute to early stages of L2 word learning due to possible variation in learning methods (e.g., picture-based vs. word-based method). Nevertheless, this study provides a first set of evidence to suggest that learning words for both novel objects and familiar objects can be similarly related to the ability to temporarily bind information across modalities in working memory. The results have implications for understanding the common cognitive processing underlying establishment of word-concept connections that might be shared by learning in L1 and L2.

In summary, the present data indicates that the capacity to form and maintain temporarily bound auditory-verbal and visual information in working memory is related to the learning/retention of phonologically-unfamiliar words both for novel objects, a context resembling L1 word learning processes, and for familiar objects, a context resembling L2 word learning processes. This extends evidence concerning the role of cross-modal working memory binding in word learning from the written language (i.e., phonological-orthographic mapping; Wang et al., 2015) to the domain of spoken language (i.e., mapping of spoken word form to lexical semantics) in L1. Our findings also have implications for understanding the impact that different contexts have on early stages of word learning, and highlight the need for further investigation of mechanisms underlying the similarities and differences in L1 and L2 word learning and their relations to different forms of working memory binding.

### **Acknowledgements**

## CROSS-MODAL WORKING MEMORY AND WORD LEARNING

This research is partially supported by the “Aim for the Top University Project” and “Center of Learning Technology for Chinese” of National Taiwan Normal University (NTNU), sponsored by the Ministry of Education, Taiwan, R.O.C. and the “International Research-Intensive Center of Excellence Program” of NTNU and Ministry of Science and Technology, Taiwan, R.O.C. under Grant no. MOST 104-2911-I-003-301. Partial support was also provided by the US National Science Foundation (BCS-1338946; BCS-1349110). We thank Nan Zhang for her assistance with running the experiment.



## CROSS-MODAL WORKING MEMORY AND WORD LEARNING

Appendix A  
Lists of Mandarin Words

Mandarin words used in learning tasks	Mandarin words used as foils (new pairs) in associative recognition retention task
茶壺(chahu, <i>teapot</i> )	黃牛(huangniu, <i>cow</i> )
碗豆(wandou, <i>pea</i> )	孔雀(kongque, <i>peacock</i> )
蠟燭(lazhu, <i>candle</i> )	熊貓(xionghmao, <i>panda</i> )
漏斗(loudou, <i>funnel</i> )	兔子(tuzi, <i>rabbit</i> )
芹菜(qincai, <i>celery</i> )	大象(daxiang, <i>elephant</i> )
刀子(daozi, <i>knife</i> )	企鵝(qie, <i>penguin</i> )
抽屜(chouti, <i>drawer</i> )	黑熊(heixiong, <i>bear</i> )
玉米(yumi, <i>corn</i> )	鱷魚(eyu, <i>crocodile</i> )
酒杯(jiubei, <i>wineglass</i> )	犀牛(xiniu, <i>rhinocero</i> )
水瓢(shuipiao, <i>bailer</i> )	斑馬(banma, <i>zebra</i> )
手套(shoutao, <i>mitten</i> )	袋鼠(daishu, <i>kangaroo</i> )
南瓜(nangua, <i>pumpkin</i> )	鸚鵡(yingwu, <i>parrot</i> )
蘿蔔(luobo, <i>carrot</i> )	公雞(gongji, <i>rooster</i> )
掃把(saoba, <i>broom</i> )	駱駝(luotuo, <i>camel</i> )
抹布(mobu, <i>mop</i> )	狐狸(huli, <i>fox</i> )
花瓶(huaping, <i>vase</i> )	駝鳥(tuoniao, <i>ostrich</i> )
蘑菇(mogu, <i>mushroom</i> )	青蛙(qingwa, <i>frog</i> )
時鐘(shizhong, <i>clock</i> )	蜥蜴(xiyi, <i>lizard</i> )
風扇(fongshan, <i>fan</i> )	小狗(xiaogou, <i>dog</i> )
烤箱(kaoxiang, <i>oven</i> )	烏龜(wugui, <i>turtle</i> )
湯杓(tangshao, <i>spoon</i> )	蝴蝶(hudie, <i>butterfly</i> )
砧板(zhenban, <i>cutting board</i> )	螃蟹(pangxie, <i>crab</i> )
鍋蓋(guogai, <i>lid</i> )	火雞(huoji, <i>turkey</i> )
紙巾(zhijin, <i>tissue</i> )	松鼠(songshu, <i>squirrel</i> )

Note. Lists of Mandarin words taken from the study of Lan, Fang, Legault, & Li (2015). Pinyin transcriptions and English translations of the Chinese words are in parentheses.

## CROSS-MODAL WORKING MEMORY AND WORD LEARNING

Appendix B  
Lists of Picture Items

Picture items used in learning tasks	Picture items used as foils (new pairs) in associative recognition retention task
broom	curtain
carrot	pistol
mop	pear
flower	lamp
mushroom	sharpener
wristwatch	sausage
kettle	screwdriver
beans	bottle-gourd
candle	spring-onion
funnell	hammar
celery	dustpan
knife	axe
non object 1	non object 6
non object 18	non object 7
non object 29	non object 8
non object 15	non object 9
non object 13	non object 10
non object 14	non object 19
non object 11	non object 22
non object 37	non object 23
non object 20	non object 26
non object 32	non object 31
non object 35	non object 34
non object 30	non object 36

Note. Lists of picture items drawn from the study of Verma and Brysbaert (2015)

## Reference

- Allen, R. J. (2015). Memory binding. In J. D. Wright (Ed.), *International Encyclopedia of the Social & Behavioural Sciences* (2 ed.): Elsevier.
- Baddeley, A. D. (1986). *Working memory*. England: Oxford University Press.
- Baddeley, A. D. (1996). Exploring the central executive. *The Quarterly Journal of Experimental Psychology: Section A*, 49(1), 5-28.
- Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, 4, 417-423.
- Baddeley, A. D. (2003). Working memory and language: an overview. *Journal of Communication Disorders*, 36(3), 189-208. doi: 10.1016/S0021-9924(03)00019-4
- Baddeley, A. D. (2012). Working Memory: Theories, Models, and Controversies. *Annual Review of Psychology*, 63(1), 1-29. doi: 10.1146/annurev-psych-120710-100422
- Baddeley, A. D., Allen, R. J., & Hitch, G. J. (2011). Binding in visual working memory: The role of the episodic buffer. *Neuropsychologia*, 49(6), 1393-1400. doi: 10.1016/j.neuropsychologia.2010.12.042
- Baddeley, A. D., Gathercole, S. E., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, 105(1), 158.
- Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. A. Bower (Ed.), *Recent Advances in Learning and Motivation* (Vol. 8). New York: Academic Press.
- Barcroft, J., & Sunderman, G. (2008). Learning new words for objects and nonobjects: Theoretical and methodological implications. *Mental Lexicon*, 3(3), 325-348. doi: 10.1075/ml.3.3.05bar
- Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General*, 133(1), 83.
- Bayley, P., Wixted, J. T., Hopkins, R. O., & Squire, L. R. (2008). Yes/no recognition, forced-choice recognition, and the human hippocampus. *Journal of Cognitive Neuroscience*, 20(3), 505-512.
- Blomert, L. (2011). The neural signature of orthographic–phonological binding in successful and failing reading development. *NeuroImage*, 57(3), 695-703. doi: 10.1016/j.neuroimage.2010.11.003
- Brockmole, J. R., & Franconeri, S. L. (2009). *Binding: A special issue of visual cognition*. Hove, UK: Psychology Press.
- Cowan, N. (1999). An embedded-processes model of working memory. *Models of working memory: Mechanisms of active maintenance and executive control*, 20, 506.
- Cowan, N. (2010). The magical mystery four: how is working memory capacity limited, and

## CROSS-MODAL WORKING MEMORY AND WORD LEARNING

- why? *Current Directions in Psychological Science*, 19(1), 51-57.
- Gathercole, S. E., Willis, C. S., Emslie, H., & Baddeley, A. D. (1992). Phonological memory and vocabulary development during the early school years: A longitudinal study. *Developmental Psychology*, 28(5), 887-898. doi: 10.1037/0012-1649.28.5.887
- Gupta, P. (2003). Examining the relationship between word learning, nonword repetition, and immediate serial recall in adults. *The Quarterly Journal of Experimental Psychology: Section A*, 56(7), 1213-1236.
- Hernandez, A., Li, P., & MacWhinney, B. (2005). The emergence of competing modules in bilingualism. *Trends in Cognitive Sciences*, 9(5), 220-225.
- Holdstock, J., Mayes, A., Roberts, N., Cezayirli, E., Isaac, C., O'reilly, R., & Norman, K. (2002). Under what conditions is recognition spared relative to recall after selective hippocampal damage in humans? *Hippocampus*, 12(3), 341-351.
- Hulme, C., Goetz, K., Gooch, D., Adams, J., & Snowling, M. J. (2007). Paired-associate learning, phoneme awareness, and learning to read. *Journal of Experimental Child Psychology*, 96(2), 150-166. doi: 10.1016/j.jecp.2006.09.002
- Jackson, O., & Schacter, D. L. (2004). Encoding activity in anterior medial temporal lobe supports subsequent associative recognition. *NeuroImage*, 21(1), 456-462.
- Jones, M. W., Branigan, H. P., Parra, M. A., & Logie, R. H. (2013). Cross-modal binding in developmental Dyslexia. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. doi: 10.1037/a0033334
- Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, 33(2), 149-174.
- Lan, Y.-J., Fang, S.-Y., Legault, J., & Li, P. (2015). Second language acquisition of Mandarin Chinese vocabulary: context of learning effects. *Educational Technology Research and Development*, 1-20.
- Li, H., Shu, H., McBride-Chang, C., Liu, H. Y., & Xue, J. (2009). Paired associate learning in Chinese children with dyslexia. *Journal of Experimental Child Psychology*, 103(2), 135-151. doi: 10.1016/j.jecp.2009.02.001
- Li, P., Zhang, F., Tsai, E., & Puls, B. (2014). Language history questionnaire (LHQ 2.0): A new dynamic web-based research tool. *Bilingualism: Language and Cognition*, 17(03), 673-680.
- Macmillan, N. A., & Kaplan, H. L. (1985). Detection theory analysis of group data: estimating sensitivity from average hit and false-alarm rates. *Psychological Bulletin*, 98(1), 185.
- Majerus, S., Poncelet, M., Greffe, C., & van der Linden, M. (2006). Relations between vocabulary development and verbal short-term memory: The relative importance of short-term memory for serial order and item information. *Journal of Experimental*

## CROSS-MODAL WORKING MEMORY AND WORD LEARNING

- Child Psychology, 93(2), 95-119.
- Poarch, G. J., van Hell, J. G., & Kroll, J. F. (2015). Accessing word meaning in beginning second language learners: Lexical or conceptual mediation? *Bilingualism: Language and Cognition*, 18(03), 357-371.
- Raven, J. C., Court, J. H., & Raven, J. (2006). *Raven's Progressive Matrices*. U.S.A.: Harcourt Assessment, Inc.
- Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, 96(4), 523-568. doi: 10.1037/0033-295X.96.4.523
- Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior research methods, instruments, & computers*, 31(1), 137-149.
- Unsworth, N., & Engle, R. W. (2007). On the division of short-term and working memory: an examination of simple and complex span and their relation to higher order abilities. *Psychological Bulletin*, 133(6), 1038.
- Verma, A., & Brysbaert, M. (2015). A validated set of tool pictures with matched objects and non-objects for laterality research. *Laterality: Asymmetries of Body, Brain and Cognition*, 20(1), 22-48. doi: 10.1080/1357650X.2014.914949
- Wang, S., Allen, R. J., Lee, J. R., & Hsieh, C.-E. (2015). Evaluating the developmental trajectory of the episodic buffer component of working memory and its relation to word recognition in children. *Journal of Experimental Child Psychology*, 133, 16-28. doi: 10.1016/j.jecp.2015.01.002
- Wang, S., & Gathercole, S. E. (2013). Working memory deficits in children with reading difficulties: Memory span and dual task coordination. *Journal of Experimental Child Psychology*, 115(1), 188-197. doi: 10.1016/j.jecp.2012.11.015
- Warmington, M., & Hulme, C. (2012). Phoneme awareness, visual-verbal paired-associate learning, and rapid automatized naming as predictors of individual differences in reading ability. *Scientific Studies of Reading*, 16(1), 45-62. doi: 10.1080/10888438.2010.534832

## Footnotes

<sup>1</sup> The results were essentially identical when we also calculated the score based on all occurrences of the target. This was also the case when  $d'$  scores were calculated based on the response to all 48 foils (24 rearranged pairs and 24 new pairs). In order to be concise, the current study only reports  $d'$  scores that were calculated based on the response to the first occurrence of 24 target pairs and on the response to the 24 rearranged pairs.

<sup>2</sup> Possible contributions of symbol recognition skills to outcomes related to binding performances can be ruled out by statistically controlling for performance on a corresponding auditory nonword task as described in the auditory-verbal condition.