Grasping Virtual Objects Benefits Lower Aptitude Learners' Acquisition of Foreign Language Vocabulary

Manuela Macedonia[1] · Brian Mathias[2] · Astrid E. Lehner[3] ·
Susanne M. Reiterer[4] · Claudia Repetto[5]

Abstract

Theories of embodied cognition describe language acquisition and representation as dependent on sensorimotor experiences that are collected during learning. Whereas native language words are typically acquired through sensorimotor experiences, foreign language (L2) words are often learned by reading, listening or repeating bilingual word lists. Recently, grasping referent objects has been demonstrated to serve as a useful sensorimotor strategy for L2 vocabulary learning. The effects of grasping virtual objects, however, remain unknown. In a virtual reality cave, we trained adult participants ($N = 46$) having low language aptitude and high language aptitude on novel L2 words under three conditions. In an audiovisual (baseline) condition, participants were presented with written and spoken L2 words. In an audiovisual observation condition, participants additionally saw virtual visual objects that corresponded to the meanings of L2 words. In an audiovisual, an observation, and a grasping condition, participants were asked to grasp the virtual objects. Participants' word learning was assessed in free and cued recall tests administered immediately after training. Relative to baseline learning, simply viewing virtual objects during learning benefitted both groups. As expected, grasping virtual objects was found to benefit vocabulary retention in low language aptitude learners. Interestingly, this benefit was not observed in high language aptitude learners. Language learning aptitude scores correlated with vocabulary learning outcomes in both audiovisual learning conditions, but not in the sensorimotor condition, suggesting that grasping altered the typical relationship between aptitude and language learning performance. The findings are interpreted in terms of differences in the extent to which procedural and declarative memory systems are accessed in low language aptitude and high language aptitude learners during sensorimotor-based vocabulary learning. Additionally, the results suggest that simulated interactions without tangible feed-

Manuela Macedonia, Brian Mathias Joint first authors.

Extended author information available on the last page of the article

back can benefit learning. This outcome expands our understanding of how physical experience extends cognition and paves the way for the application of sensorimotor enrichment strategies to virtual environments.

Introduction

According to theories of embodied and grounded cognition, language learning is rooted in bodily experiences that we collect while interacting with the world around us (Barsalou, 1999, 2008, 2020; Borghi, 2004; Jeannerod, 2006). One of the first steps in children's language acquisition is the naming of objects that can be reached, touched, dropped, tasted, and smelled. When acquiring novel words, children do not memorize the sequence of phonemes provided by their caregivers per se, but rather connect the sequence with their personal sensorimotor experiences (Glenberg & Gallese, 2012; Willems & Hagoort, 2007). These variable sensorimotor experiences may lead to differing representations of words in the brain. For example, children living in the Alps might represent the word *shark* more in visual regions of the cortex, whereas the word *trout* could be grounded in regions underlying haptic interaction (Kiefer & Pulvermüller, 2012; Pulvermüller, 1999, 2005), taste (Barrós-Loscertales et al., 2012), or smell (González et al., 2006). In the current study, we assumed that adults might also benefit from sensorimotor experiences connected with novel words in a foreign language (L2). Furthermore, we asked whether the impact of grasping virtual objects might benefit learners and the extent of these potential benefits.

Neuroscientific and behavioral research on native language (L1) processing converges in demonstrating that how we understand words is closely intertwined with sensorimotor experience, therefore refutes the idea that language is a system of abstract symbols (Fodor, 1979). Neuroscientific studies have found that the reading of motor-related words readily elicits motoric brain responses, even in the absence of overt movement (Buccino et al., 2009; Chao & Martin, 2000; Grafton et al., 1997), and these responses show a high fidelity to previous sensorimotor experiences, occurring at a level of specificity of individual limb movements (Hauk et al., 2004). Words referring to tools elicit enhanced motor activity in the brain relative to words that refer to less manipulable objects (Just et al., 2010; Rueschemeyer et al., 2010). Several behavioural lines of work have also examined the close relationship between word semantics and action. Affordance compatibility studies focus on how the perception of words and objects can influence actions executed by the hands and feet (Gibson, 1977, 1979). Response times to words in behavioural experiments are typically faster if the effector (hand or feet) is compatible with a word's semantics or even with its location (Ambrosecchia et al., 2015; Marino et al., 2013). These findings suggest that motor-related words can prime associated movements. Thus, words and the objects to which they refer may share cognitive representations and access the same motor codes (Gough et al., 2012).

In a study on novel word learning (Gordon et al., 2019), participants grasped virtual objects with either their left or their right hand, and learned their names. The participants then completed a word-colour match task. Response times were shorter for words whose response hand was the same as the hand used to grasp the object ear-

lier in the study. In a word learning study conducted by Madan and Singhal (2012), participants were asked to judge features of words such as their length and function during encoding. Concrete words that ranked highly in terms of motor manipulability such as *camera* were better memorized than words with reduced motor-related associations such as the word *table*. Taken together, these results refute the notion that words are merely abstract symbols disconnected from our everyday experience, as Fodor (1979) once suggested. Instead, words are represented through experience-related sensorimotor brain networks (Pulvermüller, 2005, 2018).

In theories of human evolution, actions are often described as the basis for language (Fischer, 2012; Rizzolatti & Arbib, 1998). Arbib (2008) proposes that movements that occurred while interacting with objects became more and more abstract and symbolic across stages of evolution and progressively evolved into gestures. Gestures, together with first vocalizations, may have formed the basis of a protolanguage. The shift from movement-based communication towards the use of spoken language could potentially have been driven by the increasing necessity to communicate abstract meanings that gestures could not sufficiently represent (Corballis, 2009a, b). Hence, from an evolutionary perspective, actions such as grasping and manipulating objects represent the underpinnings of how concepts are comprehended and words are acquired. Other theories suggest that, language may have evolved not only as a means for immediate communication, but also as a tool for triggering memory (Allen & Saidel, 1998; Hockett, 1963; Paivio, 2007). In other words, language and sensorimotor experience belong together.

Grounding Foreign Language Learning in Sensorimotor Experiences Can Benefit Learning Outcomes

Compared to learning vocabulary in one's native language, learning words in an L2 is often met with little success. Although a number of methods designed to facilitate L2 learning have been developed (Hald et al., 2016), in practice, students generally engage in listening and comprehension activities, as well as in the repetition of bilingual word lists until the meanings of L2 words have been memorized (Rasouli & Jafari, 2016). Visual-only and audiovisual strategies (e.g., reading) have been shown to be less than optimal: L2 words that are learned visually or audiovisually decay fast from memory (Barcroft, 2009; Yamamoto, 2014). One reason why the reading of written word lists may be such a popular learning strategy is because L2 instructional practice has traditionally closely followed principles of generative linguistics (Marino & Gervain, 2019). In generative linguistics, language is described as an amodal and symbolic phenomenon of the mind (Fodor, 1979). This view subscribes to the Cartesian dichotomy between body and mind, and does not consider how cognitive processes are intertwined with concurrent perceptual processing, body movements, and the physical and social environment (Barsalou, 2020). Recent work suggests that, to the contrary, integrating the body, and gestures in particular, into the learning experience, can alter how vocabulary is remembered and represented relative to passive audiovisual-only learning (Andrä et al., 2020; Macedonia, 2019; Schmidt et al., 2019). The positive impact of performing congruent gestures and actions during learning on subsequent memory performance has been referred to as

enactment effect, *subject-performed task effect*, *production effect*, and *sensorimotor enrichment benefit* (Cohen, 1981; Macedonia et al., 2011; Mathias et al., 2023; Mayer et al., 2015; Repetto et al., 2021).

Empirical research on the use of gestures in L2 learning began with Quinn-Allen's (1995) seminal work showing that emblematic gestures, i.e., culture specific gestures that convey verbal meaning such as pointing the thumbs upward, support phrase learning. Since then, a number of studies have confirmed that performing gestures while learning L2 vocabulary facilitates the memorization of the words compared to learning the words audiovisually (for reviews see Macedonia, 2014; Macedonia & von Kriegstein, 2012). Neuroscientific studies attribute the enhanced memorization of gesture-enriched words to the creation of sensorimotor brain networks associated with novel phoneme sequences. In brain imaging experiments, these networks resonate upon stimulation in a similar manner as in L1 (Macedonia et al., 2011; Mayer et al., 2015, 2017). More generally, if individuals learn words with sensorimotor input and subsequently encounter the words, sensorimotor brain regions engaged during learning also respond during the subsequent encounter, even in the absence of sensorimotor input (Fischer & Zwaan, 2008; Pulvermüller et al., 2005; Tettamanti et al., 2005). Evidence from neurostimulation studies suggests that sensory and motor areas of the cortex are causally engaged in the processing of words that have previously been learned with sensorimotor input such as gestures. In two studies, participants learned L2 words that were accompanied by either gestures or pictures. After learning, transcranial magnetic stimulation was applied to motor-related areas of the cortex, which selectively disrupted the translation of L2 words learned with gestures (Mathias et al., 2021a; Mathias et al., 2021b). This implies that the motor cortices can critically support newly-acquired word representations (Repetto et al., 2013).

Several theories aim to account for the benefits of integrating sensorimotor experiences into learning (for review see Mathias & von Kriegstein, 2023). Paivio's Dual Coding Theory (DCT), for example, describes a linguistic item as consisting of a verbal and a nonverbal code and that this dual code enhances memory for that item (Paivio & Csapo, 1969, 1973; Paivio, 1990, 2007; Sadoski & Paivio, 2012). Paivio and Desrochers (1980) apply the DCT to L2 learning specifically (see also Sadoski, 2012; Spada, 1997). Interestingly, recent theoretical approaches to embodied cognition also emphasize that the grounding of language learning can extend beyond overt physical movements. The grounding, i.e., the multiple components of information enriching a word, can also include the physical and social environment. This novel concept is referred to as 4E cognition—cognition that is embodied, embedded, enactive, and extended (Barsalou, 2020; Jusslin et al., 2022).

Language Aptitude and Vocabulary Learning

Language aptitude (LA) has been described as the capacity for an individual to achieve a higher level of ultimate attainment in a language relative to other individuals within the same time period (Carroll, 1981; Robinson, 2012). Factors such as the duration, age of onset, teaching style, or learning style of L2 learning cannot alone explain variability in language learning outcomes. In fact, differential outcomes can even be found in rather homogeneous learner groups who receive the same instruc-

tion. This observation has been taken as evidence for individual differences between language learners. Several psychological models describe the phenomenon of LA. According to Carroll (1962), LA comprises phonetic coding ability, i.e. the capacity to perceive, associate and retain sounds, inductive learning ability (capacity to induce language structure rules), grammatical sensitivity (ability to infer grammatical functions), and associative memory. Skehan (2002) proposed as capacities for LA phonetic coding (input processing), grammatical analytic ability, and memory retrieval, emphasizing the influence of working memory on each of these components. Another influential model is the Aptitude Complex Hypothesis, developed by Robinson (2001). There, the author proposes that the primary abilities underlying language learning are working memory, pattern recognition, grammatical sensitivity, and speed of processing in phonological working memory. Secondary abilities are assumed to support language learning and include memory for contingent speech, deep semantic processing, and metalinguistic rule rehearsal (for a theoretical overview, see Ameringer et al., 2018; Turker et al., 2019).

Taken per se, the learning of new vocabulary relies arguably more on memory than on other cognitive abilities. In fact, a recent study on the intentional learning of L2 Welsh vocabulary shows that short-term and working memory plays a larger role in word learning than in auditory and phonological tasks (Bisson et al., 2021). Interestingly, language instructors have traditionally not taken individual differences in memory capacity into account. This is probably due to the influence of the theory of Universal Grammar (Chomsky, 1957, 1975) in which memory was not described as a core component underlying linguistic abilities. The view that memory is a critical skill needed for the acquisition of language did not arise from linguistics. Rather, this notion came from memory research itself. Baddeley (2003) described working memory as the basis of language acquisition, as language cannot be learned without the capacity to memorize phonemes, morphemes, grammatical structure, and vocabulary. Evidence that memory is fundamental to language learning comes from studies investigating the relationship between learners' ages and their L2 learning success. The older learners are at the time of learning, the more limited are their memory capacities, and corresponding L2 learning outcomes (Hertzog et al., 2020; Whiting et al., 2011; Palmer & Havelka, 2010; Laumann Long Lisa, 2000).

Procedural and Declarative Memory for Language Learning

Traditionally, (audiovisual) word learning has been associated with capacities residing in declarative memory (Tulving & Madigan, 1970; Ullman, 2004; Brem et al., 2013). Declarative memory is typically engaged while reading, listening, or watching information (Eichenbaum, 2004). At a neural level, declarative memory is associated with hippocampal structures in the medial temporal lobe, which have been linked to word memorization (Cabeza & Moscovitch, 2013). Declarative memory capacities also correspond to increased gyrification, higher grey matter density, and greater cortical thickness in language areas, the hippocampus, and the angular gyrus, a region mediating multimodal integration (Kumar et al., 2021). Declarative memory for L2 words has also been linked to increased responses within the angular gyrus and extrastriatal cortices (Macedonia et al., 2010; Macedonia & Mueller, 2016).

There is evidence that sensorimotor-enriched L2 words rely on brain areas related not only to declarative memory such as those in the anterior temporal lobe, but also procedural memory located in the premotor and motor cortices, basal ganglia, and cerebellum, due to movement-related input that occurs during learning (Macedonia & Mueller, 2016; Pulvermüller, 2005). Procedural memory refers to memories that are encoded and retrieved in an implicit manner (Schacter, 1987). Individual differences in declarative and procedural memory circuits may underlie individual differences in storing information. For example, high-achieving verbal learners may rely primarily on declarative memory as opposed to procedural memory (Ullman & Lovelett, 2016). Nevertheless, previous work on the benefits of integrating complementary movements such as gestures into learning has not typically examined the relationship between individual differences in learning abilities and the magnitude of such benefits (Mathias & von Kriegstein, 2023).

The interplay between declarative and procedural memory in L2 learning may also depend on an individual's LA. High language aptitude learners might find success in traditional methods that heavily rely on declarative memory, while LLA learners could benefit from more sensorimotor-enriched methods that engage the procedural memory system (Macedonia et al., 2010). High language aptitude learners often possess strong declarative memory skills, which are well-suited for traditional word-learning methods (Ullman, 2004). These learners are also adept at handling higher cognitive loads. Note that cognitive load is essentially a measure of the working memory resources, according to Sweller (1988). On the other hand, individuals with LLA typically face challenges in traditional L2 learning settings due to weaker declarative memory abilities. Macedonia and Müller (2016) have shown that procedural memory systems are engaged by sensorimotor-enriched vocabulary learning tasks. It is conceivable that LLA learners might benefit from such strategies as indicated in a study with LLA learners (Macedonia et al., 2010).

Vocabulary Learning in Virtual Reality (VR) Environments

Virtual reality (VR) technology in L2 instruction presents opportunities for innovation including sensorimotor-enriched learning. Unlike traditional classroom or online methods, VR can provide an immersive experience, with multiple sensory input (Macedonia et al., 2014). Additionally, VR can allow learners the flexibility for autonomous and self-directed study (Lindgren & Johnson-Glenberg, 2013; Repetto et al., 2016). Altogether, VR has been shown to be a useful tool to enact linguistic knowledge, as described by Tuena and colleagues (2019). The utility of VR for learning seems to hinge on the extent to which individuals feel "present" in the VR-mediated environment (Johnson-Glenberg et al., 2021; Mikropoulos & Natsis, 2011). Several language learning VR studies have used avatars whose body movements are controlled by participants as a way to promote learning engagement (Chen, 2016; Ibáñez et al., 2011; Wang et al., 2017). Legault and colleagues (2019) showed that the manipulation of objects in immersive VR environments aided vocabulary learning relative to non-VR traditional word-word associative learning. In another study, Fuhrman et al. (2020) also found that manipulating objects in VR improved vocabulary memory relative to the enactment of irrelevant movements. One limitation of

these studies is that object interaction was generally conducted through button-press actions on a controller, rather than through genuine self-enactment. It is possible that sensorimotor interventions that involve VR could be more beneficial if participants are able to self-enact virtual movements. This idea aligns with an argument put forth by Johnson-Glenberg et al. (2014). In their article, they contend that the degree to which VR environments engage the motor system is a critical factor in their educational efficacy.

Aims and Hypotheses

The aim of the current study was to test whether the integration of grasping movements into language learning—a natural learning strategy present in first language acquisition—benefits adult L2 learning outcomes in LLA individuals compared to audiovisual learning. The L1 acquisition mechanism of grasping was simulated in a virtual reality (VR) cave. Adult participants were trained on L2 words in three conditions: In an audiovisual condition (AV), subjects viewed and heard L2 words; in an audiovisual and observation (AVO) condition participants viewed and heard L2 words and saw referent objects; finally, in the audiovisual, observation, and grasping (AVOG) condition, participants grasped virtual objects representing the words to memorize.

We postulated four hypotheses. First, we expected that integrating grasping movements into the learning process would, on average, benefit learning outcomes in all learners relative to audiovisual-only (AV and AVO) learning. Second, we predicted that LLA learners would benefit more from the integration of grasping movements than HLA learners. This hypothesis was based on the expectation that, if procedural learning is incorporated into the learning process, this would potentially support L2 vocabulary memorization by reducing LLA learners' cognitive load during learning (Ullman & Lovelett, 2016). Third, we tested whether language aptitude was positively associated with vocabulary learning outcomes, and fourth, whether age predicted language aptitude based on prior research showing differences in aptitude between younger and older learners (Gómez, 2017).

Methods

Participants

Forty-six participants with German as an L1 took part in the experiment (M age = 36.6 years, SD = 15.6 years, range: 19–68 years, 27 females and 19 males). A mixed effects modeling power analysis (Judd et al., 2017) based on effects of sensorimotor enrichment on L2 vocabulary learning observed by Repetto et al. (2017), an alpha level of 0.05, and power level of 0.8, suggested a minimum sample size of $N = 34$ total participants. Participants were recruited from a Linz University database, as well as through advertisements placed at the University and at the Ars Electronica Center (AEC, www.aec.at) located in Linz, Austria. All of the participants indicated that they had knowledge of at least two late-learned foreign languages (starting after

the age of 12). Self-rated L2 proficiencies ranged from low to high. No early bilinguals, individuals who regularly learned two languages before the age of 12 (Houwer, 2012), were included in the study. None of the participants reported any vision or hearing impairments, or history of neurological or psychiatric disorders. All participants provided written informed consent prior to testing. Participants received an AEC entry voucher worth €10 for their participation. The study was approved by the Ethics Committee at the University of Linz.

Materials

L1 and L2 vocabulary. The stimulus material comprised 18 words from the Vimmi language corpus, a corpus of artificial vocabulary created for studies on L2 learning in order to avoid associations with participants' native or foreign languages (Macedonia et al., 2011). Half of the words had two syllables, and the other half had three syllables. The 18 Vimmi words were associated with 18 German language translations, whose number of syllables, overall word length in letters, and frequency of use in written German (https://wortschatz.uni-leipzig.de/de) were equally distributed across experimental conditions (syllable number: $M = 2.4$, $SD = 0.9$; word length: $M = 7.4$, $SD = 2.6$; frequency: $M = 12.2$, $SD = 2.7$). The German translations were all concrete nouns referring to manipulable objects. The initial and final phonemes of the Vimmi words and their German translations always differed (see Table 1). Vimmi items were paired with concrete nouns in L1 denominating graspable objects.

Audio recordings. Vimmi and German words were recorded using a Rode NT55 microphone (Rode Microphones) in a sound-dampened chamber. An Italian native speaker recorded the Vimmi words with an Italian accent to highlight the L2 aspect

Table 1 German and Vimmi words used in the experiment, and their English translations

German word	Vimmi word	English translation
Computer	bofe	computer
Spiegel	wasute	mirror
Brille	toze	eyeglasses
Rucksack	mebeti	backpack
Gabel	bekoni	fork
Kleiderbügel	dalo	hanger
Kamera	dawu	camera
Geschenk	fapoge	present
Hammer	lefa	hammer
Flagge	nabita	flag
Seife	dotewe	soap
Korkenzieher	redu	corkscrew
Handy	igro	mobile phone
Glas	sokitu	(water) glass
Schlüssel	dupi	key
Stift	dizela	pen
Regenschirm	boruda	umbrella
Büroklammer	zobu	paper clip

of the stimuli for German-speaking participants. Vimmi audio stimuli ranged from 654 to 850 ms in length ($M = 819.7$ ms, $SD = 47.3$ ms). For more details on the audio files used in the current study, see Mayer et al. (2015).

Picture stimuli. Eighteen object pictures (Fig. 1a) corresponding to the meanings of the German words were used in the experiment. The object pictures were presented dynamically such that they "fell" from the ceiling of the VR-cave (so-called Deep Space) at the AEC into an underwater coral reef scene (Fig. 1b). The reef's underwater environment offered a plausible context for objects to appear and disappear from the same position, as if being dropped from a boat overhead. All pictures were black and white in order to exclude the influence of colour on word memoriza-
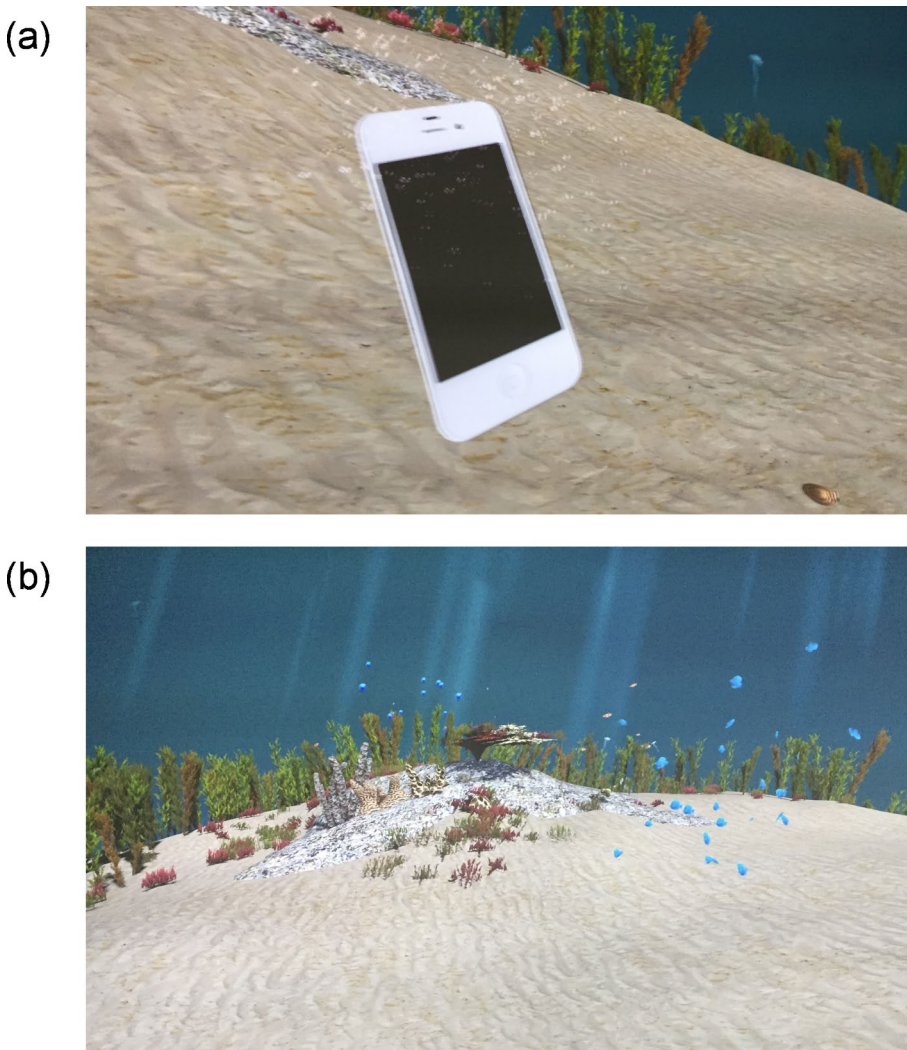
(a)

(b)

Fig. 1 **Visual stimuli used in the experiment. (a)** Virtual object plunging into underwater coral reef. **(b)** Coral reef virtual reality (VR) scene in which the virtual objects appeared

tion (Hoffmann & Engelkamp, 2017). The VR cave offers two projection spaces of 16×9 m each, one on the wall and another on the floor, with an ultra-high resolution of 8 K for stereoscopic 3D visualizations. This corresponds to a resolution of 8.192 $\times$ 4.320 pixels on each of the two projection areas, totalling more than 70 million pixels. This ultra-high definition resolution is achieved by eight Christie Boxer 4k30 Mirage 120 Hz projectors, combined with two high performance computing workstations equivalent in processing power to 400 ordinary office computers. A 5.1 Surround Sound system with Kling & Freitag speakers is used to deliver audio. Due to these unique properties of the Deep Space cave, visitors can be completely immersed into cinematic, photographic, or virtual scenes. In order to experience these scenes, 3D glasses must be worn inside the Deep Space. For our experiment, a VR learning program was developed with Unity 5.4 software (Unity Technologies, San Francisco, USA) by programmers from Johannes Kepler University Linz, Ars Electronica Solutions (www.aec.at/solutions) and Ars Electronica Future Lab (www.aec.at/futurelab). Devised as an app, the program was started by the experimenter directly from the Deep Space computer system by selecting the app from the computer screen and by starting the program with an XBOX 360 wireless controller (Microsoft Corporation, Redmond, USA).

Design

The study utilized a 2×3 mixed design with the between-participant factor aptitude (low, high) and the within-participant factor learning condition audiovisual (AV), audiovisual observation (AVO), and audiovisual observation and grasping (AVOG). The order in which the three training conditions were completed was counterbalanced across participants. The assignment of the Vimmi and German words to the learning conditions was counterbalanced across groups of participants, such that each Vimmi and German word was equally represented among the three conditions across participants.

Procedure

Vocabulary training phase. During each training trial, a written L2 (Vimmi) word and its translation into L1 (German) were projected in large yellow font on the Deep Space walls in the center of the coral reef scene for a total of 5 s. In the audiovisual (AV) condition, an audio recording of the spoken Vimmi word was presented 1 s after the written words appeared. After an inter-trial interval (ITI) of 4 s during which an empty coral reef was shown on the screen, the next trial began. In the audiovisual and observation (AVO) condition, the written L1 and L2 and spoken L2 words were accompanied by a virtual picture of the object to which the words referred. The object was presented dynamically such that it "fell" from the ceiling of the Deep Space into the water shown in the coral reef scene. The object took 1 s to fall and land on the coral reef ground, where it remained motionless for 9 s and faded away prior to the 4-s ITI. In the AVO condition, participants were instructed to simply observe the objects and not interact with them. The AVOG condition was identical to the AVO condition, except that participants were instructed to grasp the virtual objects imme-

diately after they had reached the ground and to remain grasping them until they faded away. To minimize fatigue, participants remained seated during the AV condition and stood during the AVO and AVOG conditions, in line with previous studies using similar training paradigms (e.g., Mayer et al., 2015).

Training trials were blocked by learning condition such that there were 3 total blocks. There were 72 trials in each block (6 L1-L2 word pairs × 12 repetitions). Trials were pseudo-randomly ordered within each block such that the same Vimmi word was never presented twice in a row. The AV condition lasted 10 min and the AVO and AVOG conditions each lasted 35 min due to the time that the object remained on the Deep Space walls. In a pilot experiment, the AV trials were made the same length as the AVO and AVOG trials. However, participants reported not being able to pay attention for 10-s trials during which no stimuli other than written and spoken words were presented. To facilitate attention, the long gap in stimulus presentation in the AV condition relative to the AVO and AVOG conditions was reduced. Three-min breaks occurred between each of the learning blocks during which participants were provided with water and snacks. A 5-min break followed the final training block. In total, the training lasted 80 min.

The Deep Space walls are sufficiently large to allow for the training of up to six participants simultaneously with six different object projections. We therefore trained groups of six participants simultaneously. The six participants' positions within the Deep Space were counterbalanced across learning conditions such that participants faced the front, left, or right of the Deep Space walls in different learning conditions.

Vocabulary test phase. After the training phase, participants' memory for the vocabulary was tested in a separate room. Five vocabulary tests were administered by computer. First, in an L2 free recall test, participants were instructed to type all L2 words that they could retrieve from the training. Second, participants completed an L1 free recall test. Third, in a paired free recall test, participants were instructed to write down all word pairs (rather than individual L1 or L2 words). In a fourth test, a cued L2 recall test, participants were presented with all 18 L1 words, which they were asked to translate into L2 by writing down the correct translation. Finally, in a cued L1 recall test, L2 items were translated into L1. The three free recall tests always occurred before the translation tests, to avoid priming participants' memory for the L1 and L2 words prior to completing the free recall tests. The order of L1 and L2 free recall tests was randomized across participants, and the order of the L1 and L2 cued recall tests was also randomized across participants. Participants were given a total of 5 min to complete each test. No participants exceeded the 5-min time limit for any of the tests.

Language aptitude tests. After the training phase of the experiment, participants completed parts B and D of the language independent LLAMA Language Aptitude test (Rogers et al., 2017; Granena & Long, 2013; Meara, 2005). The LLAMA B is a vocabulary learning task in which participants are asked to memorize the names of twenty fantasy cartoon figures in 2 min. The names are based on a Mesoamerican native language. Following the two-minute learning phase, participants performed a memory task in which they selected the figures corresponding to the twenty written names displayed in random order on a computer screen. The LLAMA D test examines the capacity to identify, recognize and memorize sound sequences, which is

crucial for the aptitude to learn foreign language words. LLAMA D scores have been found to predict L2 vocabulary acquisition outcomes (Hummel & French, 2016). In the LLAMA D test, participants are given 2 min to familiarize themselves with sound sequences and are then asked to select the sound sequences that would be used to spell novel auditorily presented two-syllable words. Both the LLAMA B and D tests were administered via computer and were completed in a fixed order with the LLAMA B test always occurring first.

Data Analysis

Vocabulary test scoring. Free recall and cued recall tests were scored by assigning a value of 1 for each correct response, and a value of 0 in case of incorrect response or lack of response. The total score for each test could range from 0 to 18. Scores on the five vocabulary tests (L1 free recall, L2 free recall, paired free recall, cued L1 recall, and cued L2 recall) were averaged for each participant and learning condition (AV, AVO, and AVOG), yielding a single composite test score for each participant and learning condition.

Language aptitude test scoring. Scores on the LLAMA B and Llama D questionnaires (% correct) were averaged for each participant to create a composite LLAMA test score. To group participants into LLA and HLA learners, we conducted a median split analysis. Based on the median composite LLAMA test score (Median = 38%), participants were split into two groups: the LLA group ($n = 21$, M LLAMA score = 27%, $SD = 7$%) and the HLA group ($n = 25$, M LLAMA score = 52%, $SD = 13$%).

Linear mixed effects modelling. We first inspected for outlying composite test scores based on the Interquartile Range (IQR), as suggested by Hoaglin and colleagues (1986). No participants were classified as outliers according to this procedure. We used a linear mixed effects modelling (LMM) approach to test our hypothesis that benefits of grasping on memory for L1 and L2 words would depend on participants' language learning aptitude. The model included fixed effects of aptitude (high, low) and learning condition (AV, AVO, AVOG) and a random intercept by participant. The aptitude factor was a binomial between-subjects factor. The mixed effects model used the AV condition as the reference level for the learning condition factor. The model was generated in R version 1.2.1335 using the 'lme4' package (Bates et al., 2015). Significance testing was performed using Satterthwaite's method implemented in the 'lmerTest' package, with an alpha level of $\alpha = 0.05$ (Kuznetsova et al., 2017). Post-hoc Tukey tests were conducted using the 'emmeans' package (Lenth et al., 2020). Cohen's d was computed as a measure of effect size.

Correlation analyses. To test whether language learning aptitude was associated with vocabulary learning outcomes, we correlated the composite LLAMA test scores with average vocabulary test scores for each learning condition by participant. We also examined whether age predicted language learning aptitude by correlating participants' ages with their composite LLAMA test scores.

Results

Grasping and Viewing Virtual Visual Objects During Learning Enhances Vocabulary Acquisition

We first tested the hypothesis that the performance of grasping movements in the AVOG condition would benefit learning relative to audiovisual learning that occurred in the AV and AVO conditions. The linear mixed effects model revealed that AVOG learning significantly enhanced vocabulary test performance relative to AV learning ($b = 0.99$; $p < .001$, $d = 1.18$), as shown in Fig. 2. AVO learning also significantly enhanced vocabulary test scores relative to learning in the AV condition ($b = 0.89$; $p < .001$, $d = 1.06$). For the full set of model results, see Table 2. Post-hoc tests indicated that, overall, vocabulary test performance following AVOG learning did not signifi-
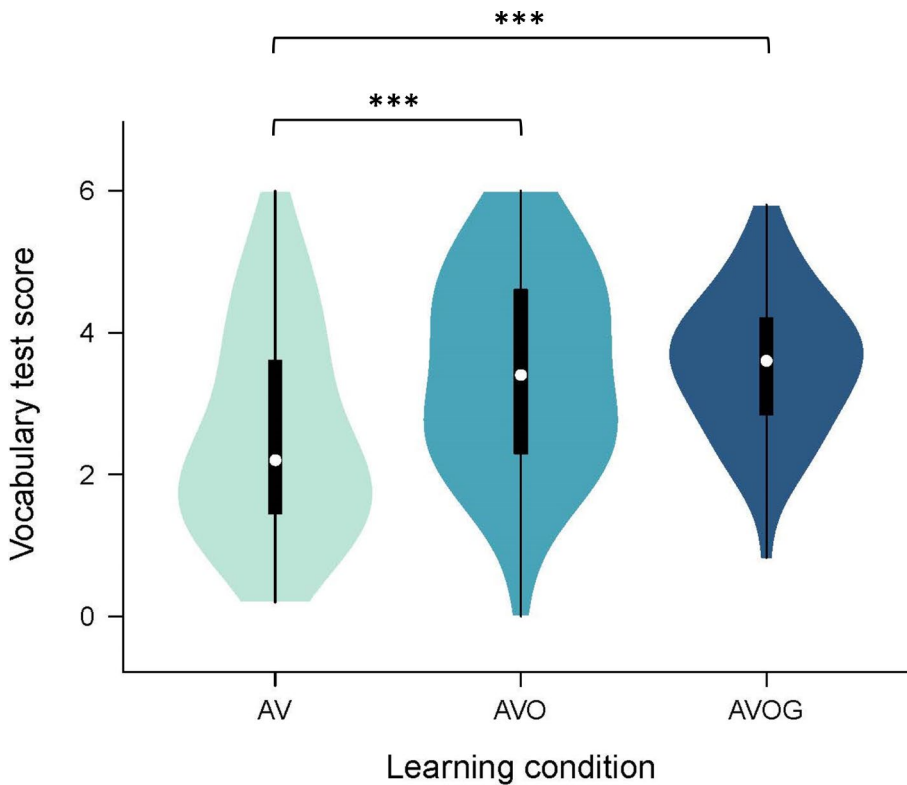
Fig. 2 Vocabulary test performance by learning condition. Viewing and grasping virtual objects referred to by foreign language (L2) words during learning significantly enhanced L2 word retention compared to viewing and hearing the L2 words. The polygons represent density estimates of the data from each condition and extend to extreme values. White circles show the medians, box limits indicate the 25th and 75th percentiles, and whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles. AV = audiovisual learning; AVO = audiovisual observation learning; AVOG = audiovisual observation and grasping learning. ***$p < .001$

Table 2 Linear mixed effects model testing the effects of learning condition and language learning aptitude on vocabulary test scores. AV = audiovisual learning; AVO = audiovisual observation learning; AVOG = audiovisual observation and grasping learning. *p < .05, ***p < .001

Fixed effects							Random effects		
	b	SE	t	p	CI	d		Variance	SD
Intercept	3.15	0.14	22.34	< 0.001	2.87, 3.43		Participant	0.67	0.82
AVO	0.89	0.18	5.08	< 0.001***	0.54, 1.24	1.06			
AVOG	0.99	0.18	5.65	< 0.001***	0.64, 1.34	1.18			
Aptitude	1.14	0.28	4.03	< 0.001***	0.57, 1.70	1.19			
AVO × Aptitude	−0.10	0.35	−0.29	0.770	−0.80, 0.59	0.06			
AVOG × Aptitude	−0.70	0.35	-2.00	0.048*	-1.40, −0.01	0.42			

Table 3 Vocabulary test scores by learning condition. AV = audiovisual learning; AVO = audiovisual observation learning; AVOG = audiovisual observation and grasping learning

	M	SD
AV	2.58	1.48
AVO	3.47	1.42
AVOG	3.54	1.02

Table 4 Vocabulary test scores by language aptitude and learning condition. LLA = low language aptitude; HLA = high language aptitude; AV = audiovisual learning; AVO = audiovisual observation learning; AVOG = audiovisual observation and grasping learning

	LLA Learners		HLA Learners	
	M	SD	M	SD
AV	1.82	1.07	3.22	1.49
AVO	2.76	1.12	4.06	1.39
AVOG	3.16	1.01	3.86	0.95

cantly exceed performance following AVO learning ($t = 0.56$, $p = .84$). Condition means and standard deviations are shown in Table 3.

Grasping Visual Objects During Learning Benefits LLA Learners More Than HLA Learners

Our second hypothesis was that integrating grasping movements into the learning process would be of greater benefit to LLA learners than to HLA learners. We therefore tested whether the aptitude factor modulated effects of learning condition on vocabulary test performance. The AVOG × Aptitude contrast was significant ($b = -0.70$; $p = .048$, $d = 0.42$), and the AVO × Aptitude contrast was not significant ($b = -0.10$; $p = .77$). This indicates that only words that had been learned in the AVOG condition (by performing grasping movements) were differently recalled depending on whether an individual participants' language learning aptitude was low or high.

Post-hoc comparisons revealed a significant difference between test scores in the AVOG and AV learning conditions for LLA learners ($t = -5.08$; $p < .001$, $d = 1.60$), but not for HLA learners ($t = -2.64$; $p = .098$), shown in Fig. 3. Thus, performing grasping movements in the AVOG learning condition did not significantly benefit HLA learners, but did significantly benefit LLA learners. Both LLA and HLA learners significantly benefited from simply observing objects in the AVO learning condition relative to the AV learning condition (LLA learners: $t = 3.57$, $p = .007$, $d = 1.12$; HLA learners: $t = 3.47$, $p = .01$, $d = 1.00$). Finally, as expected, HLA learners scored significantly higher on the vocabulary tests than LLA learners ($b = 1.14$; $p < .001$, $d = 1.19$). Condition means and standard deviations are shown in Table 4.

Language Learning Aptitude Correlates with Vocabulary Retention Following audiovisual-only Learning but Not Grasping-Based Learning

We next tested our third hypothesis regarding whether individual participants' language learning aptitudes could predict their vocabulary test scores in the three learning conditions. LLAMA scores showed a significant positive correlation with vocabulary test scores in the AV learning condition ($r(46) = 0.42$, $p = .003$) and AVO learning condition ($r(46) = 0.35$, $p = .017$), but not in the AVOG learning condition ($r(46) = 0.26$, $p = .076$), shown in Fig. 4. Thus, the addition of grasping movements to the vocabulary learning task changed the relationship between language learning aptitude and vocabulary retention. The correlation between age and language learning aptitude did not reach significance ($r(46) = -0.27$, $p = .070$), although it was in the expected direction (higher aptitude scores for younger participants).
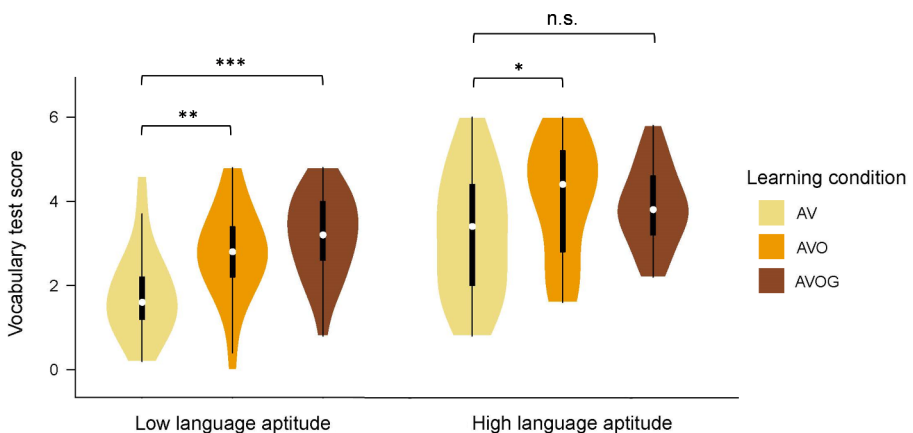

Fig. 3 Vocabulary test performance by learning condition and language learning aptitude. The integration of grasping movements into foreign language (L2) training significantly benefitted low aptitude learners, but not high aptitude learners, relative to simply viewing and hearing L2 words. The polygons represent density estimates of the data from each condition and extend to extreme values. White circles show the medians, box limits indicate the 25th and 75th percentiles, and whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles. AV = audiovisual learning; AVO = audiovisual observation learning; AVOG = audiovisual observation and grasping learning. *$p < .05$, **$p < .01$, ***$p < .001$
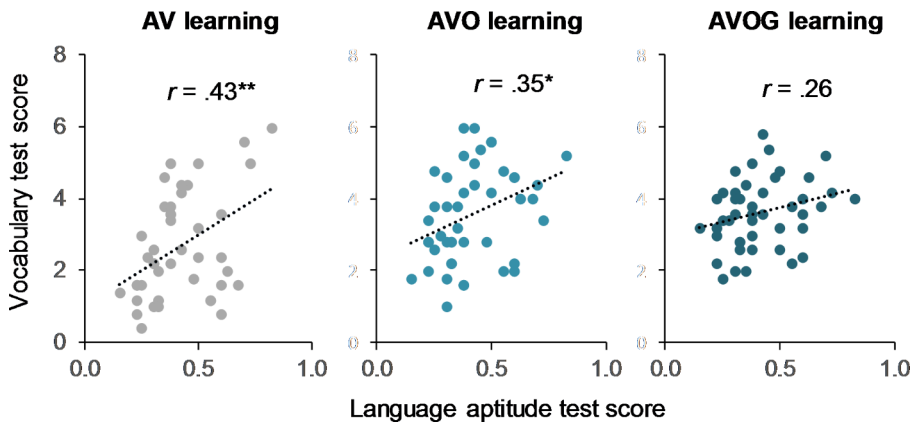
Fig. 4 **Correlations of language aptitude scores and vocabulary test scores by learning condition.** Language aptitude significantly correlated with vocabulary learning outcomes following audiovisual forms of learning, but not learning that involved grasping virtual objects. AV = audiovisual learning; AVO = audiovisual observation learning; AVOG = audiovisual observation and grasping learning. *p < .05, **p < .01

Discussion

The current study investigated whether benefits of sensorimotor-enriched L2 learning by means of grasping might differ across levels of word-learning aptitude, or whether learners of different aptitude respond similarly to sensorimotor learning interventions. Adult LLA and HLA word-learners were trained on novel L2 vocabulary by simply viewing the L2 words and their L1 translations (AV condition), viewing the L2 and L1 words along with a virtual visual object depicting the word's referent (AVO condition), and viewing the words while grasping the virtual object (AVOG condition).

In line with our first hypothesis, grasping virtual objects during learning benefitted vocabulary acquisition relative to simply viewing L2 and L1 words. However, this benefit was specific to LLA learners. HLA learners did not benefit from integrating grasping movements into learning. This result confirms our second hypothesis. Interestingly, viewing objects referred to by L2 words during L2 learning enhanced retention relative to viewing the L2 words themselves, both in LLA and HLA learners. When participants grasped the virtual objects while learning, the relationship between language learning aptitude and vocabulary learning was altered such that aptitude could no longer positively predict learning outcomes. Finally, age was not found to predict language learning aptitude in this study, contrary to our fourth hypothesis.

We interpret these findings in terms of procedural and declarative memory. Procedural memory is likely engaged during movement-enriched learning, as demonstrated by Macedonia and Müller (2016). We reason that HLA learners may inherently rely on their declarative memory for L2 learning, as long as the L2 input does not constitute too high a cognitive load. Conversely, LLA learners' word learning may be supported to a greater extent by the recruitment of procedural memory systems. In other

words, HLA learners may effectively pick up new words without needing to involve the motor system, while LLA learners' performance is improved by integrating sensorimotor elements into the learning experience.

Viewing Visual Objects during Vocabulary Learning Benefits both low and high Aptitude Learners

Viewing a virtual object during L2 learning enhanced subsequent memory performance for both low and high language aptitude learners. This finding is consistent with several recent studies that demonstrated beneficial effects of presenting complementary visual information along with written or spoken words during L2 word acquisition (Andrä et al., 2020; Mathias et al., 2022; Mayer et al., 2015). The enrichment of L2 learning with pictures has also been used as a teaching strategy in educational practice for decades, although such teaching methods (Riesenberg et al., 2009) have not been scientifically investigated until recently. The picture benefit here is consistent with cognitive and neural theories emphasizing multimodal interactions in sensorimotor enrichment learning benefits such as the DCT (Paivio & Csapo, 1969, 1973) and multisensory learning theory (Mayer et al., 2015; Mathias & von Kriegstein, 2023).

Differential Effects of Grasping Objects for low and high Aptitude Language Learners

Grasping virtual objects during word learning also enhanced word retention relative to baseline learning, but only in LLA learners and not in HLA learners. The grasping benefit demonstrated in LLA learners is consistent with numerous studies showing positive effects of sensorimotor enrichment of words and phrases by means of congruent gestures or movements (Bäckman & Nilsson, 1985; Engelkamp & Krumnacker, 1980; Engelkamp, 1980; Engelkamp et al., 1994; Engelkamp et al., 1995; Kormi-Nouri et al., 1994; Mimura et al., 1998; Zimmer, 1996; Zimmer & Saathoff, 1997). Although the current study did not involve the enactment of gestures, these findings extend the gesture results and demonstrate that object-directed movements performed while grasping virtual objects can also support L2 word retention.

The finding that grasping virtual objects—without the tactile experience of actually touching them, which was the case in the VR environment—can improve word retention in certain learners has important implications for theories of embodiment. Specifically, it challenges the idea that physical touch or tactile feedback is an essential component for activating the sensorimotor systems that facilitate learning. Instead, this research suggests that even simulated, non-tactile interactions can sufficiently engage these systems, broadening our understanding of what embodiment can entail. We therefore view this finding as supporting the grounded cognition view that cognitive processes extend beyond body movements and into the surrounding environment (Barsalou, 2020). This finding also adds a fascinating layer to our understanding of the efficacy of sensorimotor learning strategies in the context of VR environments. The fact that full tactile engagement may not always be necessary to benefit from embodied learning strategies expands the range of effective, low-

cost educational VR interventions that can be developed. The findings open up new avenues for exploring the minimum requirements for effective sensorimotor enrichment in VR environments.

In general, learning outcomes are influenced by individual variations in learning aptitude (Dahlen & Caldwell–Harris, 2013). However, pinpointing the specific mechanisms involved has proven to be a complex task. Prior studies, such as one by Poschner (2018), indicate that LLA and HLA learners don't necessarily employ different cognitive strategies for language acquisition; for instance, both groups use sound associations between foreign language (L2) words and their native language (L1) translations. Further complicating the picture, Matusz and colleagues (2017) demonstrated that HLA learners excel in integrating multisensory events within the intraparietal cortex, a neural hub crucial for selective attention, as noted by Fiebelkorn and Kastner (2020), and multisensory processing. Moreover, HLA learners display heightened activity in the left angular gyrus and right extra-striate cortex when recognizing gesture-associated words compared to LLA learners (Macedonia et al., 2010). These brain regions are also implicated in the integration of information across diverse sensory modalities (Binder et al., 2009; Seghier, 2012). Taken together, our findings suggest that LLA and HLA learners may differentially process multisensory and sensorimotor cues during their learning experiences. Note however, that the differences in accessing memory resources, i.e., the recruitment of procedural memory in LLA learners, does not occur intentionally. It seems to be an innate strategy that is applied in order to perform the task.

If HLA learners exhibit more efficient multisensory integration processes compared to LLA learners, one might expect that enrichment cues would be especially beneficial for the HLA group. This is not what we observed. Instead, the integration of grasping movements into the learning process was more beneficial for LLA than HLA learners. Why could this be so? We propose that LLA and HLA learners may differ in how declarative and procedural memory systems are deployed during vocabulary learning. The LLA and HLA learners in our study likely differed in terms of declarative memory ability, as assessed by the LLAMA B and D subtests. Declarative memory abilities have previously been associated with individual differences in phonological processing, which encompasses both phonological working memory and retrieval (Arthur et al., 2021; Baddeley, 2010) and was a core component of the current word-learning tasks. Thus, both the language aptitude tests and the audiovisual learning conditions likely engaged declarative memory. However, procedural memory was likely recruited by the grasping movements performed during learning (Macedonia & Mueller, 2016). It is possible that LLA learners relied to a greater extent than HLA learners on procedural memory systems when learning words by grasping their referent objects. This led to greater benefits of grasping for LLA learners than HLA learners relative to baseline audiovisual learning.

One possible explanation for the greater engagement of procedural memory among LLA learners compared to HLA learners could be that the LLA group leveraged procedural memory to mitigate cognitive load, economizing the working memory resources required for effective learning. Though limited in capacity (Miller, 1956; Cowan, 2001), there is some evidence that working memory can be improved by factors such as training, expertise, or even encoding strategy (Ericsson & Kintsch, 1995).

If physical actions performed during learning are congruent with to-be-learned stimuli, then these actions typically enhance task performance (Cook et al., 2008; Skulmowski & Rey, 2018). Hence, actions may be able to reduce cognitive load and serve as a successful – "natural" – strategy for LLA learners. This explanation is supported by Paas and Sweller's (2012) biological evolution theory that considers sensorimotor experiences to be sources of biologically *primary* information. While interacting with the world, individuals acquire knowledge schemas that are necessary in order to build up *secondary* biological information such as language. More importantly, by constructing schemas through sensorimotor experiences, individuals are able to save cognitive resources. The ability to gesture in order to reduce cognitive load and to externalize thoughts during speech production and language development has been also addressed by Goldin-Meadow (2001) and Ping & Goldin-Meadow (2010). We propose that the grasping movements performed in the current study saved cognitive resources via cognitive offloading to a greater extent in the LLA learners, who were defined in our study based on tests of short-term and working memory (cf. Risko & Gilbert, 2016). This cognitive offloading may, like the use of procedural memory systems, have improved the retention of sensorimotor-enriched words.

Potential Effects of Stimulus Timing and Study Limitations

It is worth noting that the lengths of trials in which participants grasped and viewed objects in the current study differed from the length of trials in which participants merely viewed and heard L2 words without seeing any objects (baseline learning trials). Object grasping and object viewing trials were roughly twice as long as trials in which no objects were presented. Despite this difference, HLA learners showed no learning advantage for grasping-enriched trials relative to baseline. Thus, the differences in trial timing between baseline and grasping conditions is not able to explain the divergence between LLA and HLA learners in terms of grasping benefits. Previous L2 learning studies that shortened the length of baseline trials (e.g., Andrä et al., 2020) and studies that have used equivalent trial lengths for all learning conditions (e.g., Macedonia et al., 2011; Mayer et al., 2015) did not observe any systematic relationship between trial lengths and vocabulary learning outcomes.

A limitation of the present study is that our findings are specifically applicable to the learning of L2 concrete words with meanings already well-understood in the learner's L1. It is conceivable that the benefits of sensorimotor enriched learning could be even more pronounced when applied to vocabulary items that are unfamiliar in both the learner's L1 and L2. In such cases, sensorimotor cues could offer crucial support for establishing entirely new semantic representations. An additional limitation concerns the nature of the gestures involved. Participants engaged in simple grasping movements, without performing more complex, functionally relevant manipulations of the objects—such as inserting and turning a key or hammering a nail. It is conceivable that functional manipulations might engage the motor system more deeply, thereby enhancing an item's distinctiveness in procedural memory. Future research could investigate the potentially larger benefits of performing more functionally meaningful gestures in VR environments on vocabulary learning outcomes.

Conclusion and Pedagogical Outlook

A growing body of research has shown that the use of sensorimotor enrichment strategies during L2 word learning can enhance the memorization of those words. The present study has demonstrated that grasping virtual objects also benefits retention, particularly in LLA learners. We propose that grasping virtual objects during learning engaged LLA learners' procedural memory. This in turn enhanced their vocabulary acquisition compared to HLA learners who benefitted from higher declarative memory capacities. Although VR is not a new technology, research on the use of VR for language learning in pedagogical settings, is rare. With the advent of VR devices that can be purchased at reasonable price, vocabulary learning with VR objects could support LLA learners in an efficient way: VR would allow training to be provided ubiquitously, accessible to everyone and at any time in facilitation of multilingualism. The technology could at the same time allow personalized programs that take into account a learner's aptitude and individual learning needs (Macedonia et al., 2014).

References

Allen, L. Q. (1995). The effects of emblematic gestures on the Development and Access of Mental representations of French expressions. *The Modern Language Journal*, *79*(4), 521–529. https://doi.org/10.1111/j.1540-4781.1995.tb05454.x.

Allen, C., & Saidel, E. (1998). The evolution of reference. In D. D. Cummins, & C. Allen (Eds.), *The evolution of mind* (pp. 184–203). Oxford University Press (OUP).

Ambrosecchia, M., Marino, B. F. M., Gawryszewski, L. G., & Riggio, L. (2015). Spatial stimulus-response compatibility and affordance effects are not ruled by the same mechanisms. *Frontiers in Human Neuroscience*, *9*(MAY), https://doi.org/10.3389/fnhum.2015.00283.

Ameringer, V., Green, L., Leisser, D., & Turker, S. (2018). *Introduction: Towards an Interdisciplinary understanding of Language Aptitude* (pp. 1–15). Springer. https://doi.org/10.1007/978-3-319-91917-1_1.

Andrä, C., Mathias, B., Schwager, A., Macedonia, M., & von Kriegstein, K. (2020). Learning Foreign Language Vocabulary with gestures and pictures enhances Vocabulary Memory for several months post-learning in eight-year-Old School Children. *Educational Psychology Review*, *32*(3), 815–850. https://doi.org/10.1007/s10648-020-09527-z.

Arbib, M. A. (2008). From grasp to language: Embodied concepts and the challenge of abstraction. *Journal of Physiology-Paris*, *102*(1–3), 4–20. https://doi.org/10.1016/j.jphysparis.2008.03.001.

Arthur, D. T., Ullman, M. T., & Earle, F. S. (2021). Declarative Memory Predicts Phonological Processing Abilities in Adulthood. In *Frontiers in Psychology* (Vol. 12, p. 1813).

Bäckman, L., & Nilsson, L. G. (1985). Prerequisites for lack of age differences in memory performance. *Experimental Aging Research*, *11*(2), 67–73. https://doi.org/10.1080/03610738508259282.

Baddeley, A. (2003). Working memory and language: An overview. *Journal of Communication Disorders*, *36*(3), 189–208. https://doi.org/10.1016/S0021-9924(03)00019-4.

Baddeley, A. (2010). Working memory. *Current Biology, 20*(4), R136–R140. https://doi.org/10.1016/j.cub.2009.12.014.

Barcroft, J. (2009). Strategies and performance in intentional L2 vocabulary learning. *Language Awareness, 18*(1), 74–89. https://doi.org/10.1080/09658410802557535.

Barrós-Loscertales, A., González, J., Pulvermüller, F., Ventura-Campos, N., Bustamante, J. C., Costumero, V., Parcet, M. A., & Ávila, C. (2012). Reading salt activates gustatory brain regions: fMRI evidence for semantic grounding in a novel sensory modality. *Cerebral Cortex (New York, N.Y.: 1991)*, *22*(11), 2554–2563. https://doi.org/10.1093/cercor/bhr324.

Barsalou, L. W. (1999). Perceptual symbol systems. *The Behavioral and Brain Sciences*, *22*(4), 577–609. discussion 610 – 60.

Barsalou, L. W. (2008). Grounded Cognition. *Annual Review of Psychology*, *59*(1), 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639.

Barsalou, L. W. (2020). Challenges and opportunities for Grounding Cognition. *Journal of Cognition*. https://doi.org/10.5334/joc.116.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01. SE-Articles.

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the Semantic System? A critical review and Meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*, *19*(12), 2767–2796. https://doi.org/10.1093/cercor/bhp055.

Bisson, M. J., Kukona, A., & Lengeris, A. (2021). An ear and eye for language: Mechanisms underlying second language word learning. *Bilingualism*, *24*(3), 549–568. https://doi.org/10.1017/S1366728920000723.

Borghi, A. M. (2004). Object concepts and action: Extracting affordances from objects parts. *Acta Psychologica*, *115*(1), 69–96. https://doi.org/10.1016/j.actpsy.2003.11.004.

Brem, A. K., Ran, K., & Pascual-Leone, A. (2013). Learning and memory. *Handbook of Clinical Neurology, 116*, 693–737. https://doi.org/10.1016/B978-0-444-53497-2.00055-3.

Buccino, G., Sato, M., Cattaneo, L., Rodà, F., & Riggio, L. (2009). Broken affordances, broken objects: A TMS study. *Neuropsychologia*, *47*(14), 3074–3078. https://doi.org/10.1016/j.neuropsychologia.2009.07.003.

Cabeza, R., & Moscovitch, M. (2013). Memory Systems, Processing modes, and components: Functional neuroimaging evidence. *Perspectives on Psychological Science*, *8*(1), 49–55. https://doi.org/10.1177/1745691612469033.

Carroll, B. J. (1981). How to develop communicative language tests. *World Englishes, 1*(1), 35–38. https://doi.org/10.1111/j.1467-971X.1981.tb00446.x.

Carroll, J. B. (1962). The prediction ofsuccess in intensive foreign language training. In R. Glaser (Ed.), *Training research and education* (pp. 87–136). University of Pittsburgh Press.

Chao, L. L., & Martin, A. (2000). Representation of manipulable man-made objects in the dorsal stream. *Neuroimage*, *12*(4), 478–484. https://doi.org/10.1006/nimg.2000.0635.

Chen, J. C. (2016). The crossroads of English language learners, task-based instruction, and 3D multi-user virtual learning in Second Life. *Computers & Education*, *102*, 152–171. https://doi.org/10.1016/j.compedu.2016.08.004.

Chomsky, N. (1957). *Syntactic Structures*. Mouton.

Chomsky, N. (1975). *Reflections on language*. Pantheon Books.

Cohen, R. L. (1981). On the generality of some memory laws. *Scandinavian Journal of Psychology, 22*(1), 267–281. https://doi.org/10.1111/j.1467-9450.1981.tb00402.x.

Cook, S. W., Mitchell, Z., & Goldin-Meadow, S. (2008). Gesturing makes learning last. *Cognition, 106*(2), 1047–1058. https://doi.org/10.1016/J.Cognition.2007.04.010.

Corballis, M. C. (2009a). The evolution of Language. *Annals of the New York Academy of Sciences, 1156*(1), 19–43. https://doi.org/10.1111/j.1749-6632.2009.04423.x.

Corballis, M. C. (2009b). Language as gesture. *Human Movement Science, 28*(5), 556–565. https://doi.org/10.1016/j.humov.2009.07.003.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*(1), 87–114. https://doi.org/10.1017/S0140525X01003922.

Dahlen, K., & Caldwell–Harris, C. (2013). Rehearsal and aptitude in foreign vocabulary learning. *The Modern Language Journal, 97*(4), 902–916. https://doi.org/10.1111/j.1540-4781.2013.12045.x.

Eichenbaum, H. (2004). Hippocampus: Cognitive processes and neural representations that underlie declarative memory. *Neuron, 44*(1), 109–120. https://doi.org/10.1016/j.neuron.2004.08.028.

Engelkamp, J. (1980). Imaginale und motorische Prozesse beim Behalten verbalen Materials. *Zeitschrift Für Experimentelle Und Angewandte Psychologie, 27*, 511–533.

Engelkamp, J., & Krumnacker, H. (1980). Image-and motor-processes in the retention of verbal materials. *Zeitschrift für experimentelle und angewandte Psychologie, 27*, 511–533.

Engelkamp, J., Zimmer, H. D., Mohr, G., & Sellen, O. (1994). Memory of self-performed tasks: Self-performing during recognition. *Memory & Cognition, 22*(1), 34–39.

Engelkamp, J., Zimmer, H. D., & Kurbjuweit, A. (1995). Verb frequency and enactment in implicit and explicit memory. *Psychological Research Psychologische Forschung, 57*(3), 242–249. https://doi.org/10.1007/BF00431285.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102*(2), 211–245. https://doi.org/10.1037/0033-295X.102.2.211.

Fiebelkorn, I. C., & Kastner, S. (2020). Functional specialization in the attention network. *Annual Review of Psychology, 71*, 221–249. https://doi.org/10.1146/annurev-psych-010418-103429.

Fischer, J. L. (2012). Grasping and the gesture theory of language origins. In Marge E. Landsberg (Ed.), *The Genesis of Language* (pp. 67–78). De Gruyter Mouton. https://doi.org/10.1515/9783110847536.67.

Fischer, M. H., & Zwaan, R. A. (2008). Embodied Language: A review of the role of the Motor System in Language Comprehension. *Quarterly Journal of Experimental Psychology, 61*(6), 825–850. https://doi.org/10.1080/17470210701623605.

Fodor, J. A. (1979). *The language of thought*. Harvard University Press.

Fuhrman, O., Eckerling, A., Friedmann, N., Tarrasch, R., & Raz, G. (2020). The moving learner: Object manipulation in virtual reality improves vocabulary learning. *Journal of Computer Assisted Learning, 37*. https://doi.org/10.1111/jcal.12515.

Gibson, J. (1977). A note on perceiving, acting, and knowing: Toward an ecological psychology by J. In R. Shaw, & J. Bransford (Eds.), *Perceiving, acting, and knowing: Toward an ecological psychology* (pp. 67–82). Erlbaum.

Gibson, J. (1979). *The Ecological Approach to Visual Perception*. Houghton-Mifflin.

Glenberg, A. M., & Gallese, V. (2012). Action-based language: A theory of language acquisition, comprehension, and production. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior, 48*(7), 905–922. https://doi.org/10.1016/J.CORTEX.2011.04.010.

Gómez, R. L. (2017). Do infants retain the statistics of a statistical learning experience? Insights from a developmental cognitive neuroscience perspective. *Philosophical Transactions of the Royal Society B: Biological Sciences, 372*(1711), 20160054. https://doi.org/10.1098/rstb.2016.0054.

Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., & Wagner, S. (2001). Explaining Math: Gesturing Lightens the load. *Psychological Science, 12*(6), 516–522. https://doi.org/10.1111/1467-9280.00395.

González, J., Barros-Loscertales, A., Pulvermüller, F., Meseguer, V., Sanjuán, A., Belloch, V., & Ávila, C. (2006). Reading cinnamon activates olfactory brain regions. *Neuroimage, 32*(2), 906–912. https://doi.org/10.1016/j.neuroimage.2006.03.037.

Gordon, C. L., Shea, T. M., Noelle, D. C., & Balasubramaniam, R. (2019). Affordance Compatibility Effect for Word Learning in virtual reality. *Cognitive Science, 43*(6), https://doi.org/10.1111/cogs.12742.

Gough, P. M., Riggio, L., Chersi, F., Sato, M., Fogassi, L., & Buccino, G. (2012). Nouns referring to tools and natural objects differentially modulate the motor system. *Neuropsychologia, 50*(1), 19–25. https://doi.org/10.1016/j.neuropsychologia.2011.10.017.

Grafton, S. T., Fadiga, L., Arbib, M. A., & Rizzolatti, G. (1997). Premotor cortex activation during observation and naming of familiar tools. *Neuroimage, 6*(4), 231–236. https://doi.org/10.1006/nimg.1997.0293.

Granena, G., & Long, M. H. (2013). Age of onset, length of residence, language aptitude, and ultimate L2 attainment in three linguistic domains. *Second Language Research*, *29*(3), 311–343. https://doi.org/10.1177/0267658312461497.

Hald, L. A., de Nooijer, J., van Gog, T., & Bekkering, H. (2016). Optimizing Word Learning via Links to Perceptual and Motoric Experience. *Educational Psychology Review*, *28*(3), 495–522. https://doi.org/10.1007/s10648-015-9334-2.

Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic Representation of Action Words in Human Motor and Premotor Cortex. *Neuron*, *41*(2), 301–307. https://doi.org/10.1016/S0896-6273(03)00838-9.

Hertzog, C., Price, J., & Murray, R. (2020). Age differences in item selection behaviors and subsequent memory for new foreign language vocabulary: Evidence for a region of proximal learning heuristic. *Psychology and Aging*, *35*(8), 1059–1072. https://doi.org/10.1037/PAG0000574.

Hoaglin, D. C., Iglewicz, B., & Tukey, J. W. (1986). Performance of some resistant rules for outlier labeling. *Journal of the American Statistical Association*, *81*(396), 991–999. https://doi.org/10.1080/01621459.1986.10478363.

Hockett, C. F. (1963). The problem of universals in language. *Universals of Language*, *2*, 1–29.

Hoffmann, J., & Engelkamp, J. (2017). Lern- und Gedächtnispsychologie. Springer. https://doi.org/10.1007/978-3-662-49068-6.

Houwer, A., & De (2012). Early Bilingualism. In *The Encyclopedia of Applied Linguistics*. https://doi.org/10.1002/9781405198431.wbeal0351.

Hummel, K. M., & French, L. M. (2016). Phonological memory and aptitude components: Contributions to second language proficiency. *Learning and Individual Differences*, *51*(C), 249–255. https://doi.org/10.1016/j.lindif.2016.08.016.

Ibáñez, M. B., García, J. J., Galán, S., Maroto, D., Morillo, D., & Kloos, C. D. (2011). Design and implementation of a 3D multi-user virtual world for Language Learning. *Journal of Educational Technology & Society*, *14*(4), 2–10.

Jeannerod, M. (2006). *Motor cognition: What actions tell the self*. Oxford University Press.

Johnson-Glenberg, M., Birchfield, D., Tolentino, L., & Koziupa, T. (2014). Collaborative embodied learning in mixed reality motion-capture environments: Two Science studies. *Journal of Educational Psychology*, *106*, 86. https://doi.org/10.1037/a0034008.

Johnson-Glenberg, M. C., Bartolomea, H., & Kalina, E. (2021). Platform is not destiny: Embodied learning effects comparing 2D desktop to 3D virtual reality STEM experiences. *Journal of Computer Assisted Learning*, *37*(5), 1263–1284. https://doi.org/10.1111/jcal.12567.

Judd, C. M., Westfall, J., & Kenny, D. A. (2017). Experiments with more than one random factor: Designs, analytic models, and statistical power. *Annual Review of Psychology, 68*, 601–625. https://doi.org/10.1146/annurev-psych-122414-033702.

Jusslin, S., Korpinen, K., Lilja, N., Martin, R., Lehtinen-Schnabel, J., & Anttila, E. (2022). Embodied learning and teaching approaches in language education: A mixed studies review. *Educational Research Review*, *37*, 100480. https://doi.org/10.1016/j.edurev.2022.100480.

Just, M. A., Cherkassky, V. L., Aryal, S., & Mitchell, T. M. (2010). A neurosemantic theory of concrete noun representation based on the underlying Brain codes. *Plos One*, *5*(1), e8622. https://doi.org/10.1371/journal.pone.0008622.

Kiefer, M., & Pulvermüller, F. (2012). Conceptual representations in mind and brain: Theoretical developments, current evidence and future directions. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *48*(7), 805–825. https://doi.org/10.1016/j.cortex.2011.04.006.

Kormi-Nouri, R., Nyberg, L., & Nilsson, L. G. (1994). The effect of retrieval enactment on recall of subject-performed tasks and verbal tasks. *Memory & Cognition*, *22*(6), 723–728. https://doi.org/10.3758/BF03209257.

Kumar, U., Singh, A., & Paddakanya, P. (2021). Extensive long-term verbal memory training is associated with brain plasticity. *Scientific Reports*, *11*(1), 1–12. https://doi.org/10.1038/s41598-021-89248-7.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13 SE-Articles), 1–26. https://doi.org/10.18637/jss.v082.i13.

Laumann, L., & Lisa, R. J. S. (2000). Adult age differences in vocabulary acquisition. *Educational Gerontology*, *26*(7), 651–664. https://doi.org/10.1080/03601270050200644.

Legault, J., Zhao, J., Chi, Y. A., Chen, W., Klippel, A., & Li, P. (2019). Immersive Virtual Reality as an Effective Tool for Second Language Vocabulary Learning. In *Languages* (Vol. 4, Issue 1). https://doi.org/10.3390/languages4010013.

Lenth, R. V. (2020). emmeans: Estimated Marginal Means, aka Least-Squares Means. *R Package Version 1.5.2-1*

Lindgren, R., & Johnson-Glenberg, M. (2013). Emboldened by Embodiment. *Educational Researcher*, *42*(8), 445–452. https://doi.org/10.3102/0013189X13511661.

Macedonia, M. (2014). Bringing back the body into the mind: Gestures enhance word learning in foreign language. *Frontiers in Psychology*, *5*(DEC). https://doi.org/10.3389/fpsyg.2014.01467.

Macedonia, M. (2019). Embodied Learning: Why at School the Mind Needs the Body. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.02098.

Macedonia, M., & Mueller, K. (2016). Exploring the neural representation of novel words learned through enactment in a word recognition task. *Frontiers in Psychology*, *7*(JUN). https://doi.org/10.3389/fpsyg.2016.00953.

Macedonia, M., & von Kriegstein, K. (2012). Gestures enhance Foreign Language Learning. *BIOLINGUISTICS*, *6*(3–4), 393–416.

Macedonia, M., Müller, K., & Friederici, A. D. (2010). Neural correlates of high performance in foreign language vocabulary learning. *Mind Brain and Education*, *4*(3), https://doi.org/10.1111/j.1751-228X.2010.01091.x.

Macedonia, M., Müller, K., & Friederici, A. D. (2011). The impact of iconic gestures on foreign language word learning and its neural substrate. *Human Brain Mapping*, *32*(6), 982–998. https://doi.org/10.1002/hbm.21084.

Macedonia, M., Groher, I., & Roithmayr, F. (2014). Intelligent virtual agents as language trainers facilitate multilingualism. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00295.

Madan, C. R., & Singhal, A. (2012). Encoding the world around us: Motor-related processing influences verbal memory. *Consciousness and Cognition*, *21*(3), 1563–1570. https://doi.org/10.1016/J.CONCOG.2012.07.006.

Marino, C., & Gervain, J. (2019). The impact of generative linguistics on psychology. *Acta Linguistica Academica*, *66*(3), 371–396.

Marino, B. F. M., Gough, P. M., Gallese, V., Riggio, L., & Buccino, G. (2013). How the motor system handles nouns: A behavioral study. *Psychological Research Psychologische Forschung*, *77*(1), 64–73. https://doi.org/10.1007/s00426-011-0371-2.

Mathias, B., & von Kriegstein, K. (2023). Enriched learning: Behavior, brain, and computation. *Trends in Cognitive Sciences*, *27*(1), 81–97. https://doi.org/10.1016/j.tics.2022.10.007.

Mathias, B., Andrä, C., Schwager, A., Macedonia, M., & von Kriegstein, K. (2022). Twelve-and fourteen-year-old school children differentially benefit from sensorimotor-and multisensory-enriched vocabulary training. *Educational Psychology Review*, *34*(3), 1739–1770. https://doi.org/10.1007/s10648-021-09648-z.

Mathias, B., Waibel, A., Hartwigsen, G., Sureth, L., Macedonia, M., Mayer, K. M., & von Kriegstein, K. (2021a). Motor cortex causally contributes to vocabulary translation following sensorimotor-enriched training. *Journal of Neuroscience*, *41*(41), 8618–8631. https://doi.org/10.1523/JNEUROSCI.2249-20.2021

Mathias, B., Sureth, L., Hartwigsen, G., Macedonia, M., Mayer, K. M., & von Kriegstein, K. (2021b). Visual sensory cortices causally contribute to auditory word recognition following sensorimotor-enriched vocabulary training. *Cerebral Cortex*, *31*(1), 513–528. https://doi.org/10.1093/cercor/bhaa240.

Matusz, P. J., Wallace, M. T., & Murray, M. M. (2017). A multisensory perspective on object memory. *Neuropsychologia*, *105*, 243–252. https://doi.org/10.1016/j.neuropsychologia.2017.04.008.

Mayer, K. M., Macedonia, M., & von Kriegstein, K. (2017). Recently learned foreign abstract and concrete nouns are represented in distinct cortical networks similar to the native language. *Human Brain Mapping*, *38*(9), https://doi.org/10.1002/hbm.23668.

Mayer, K. M., Yildiz, I. B., Macedonia, M., & Von Kriegstein, K. (2015). Visual and motor cortices differentially support the translation of foreign language words. *Current Biology*, *25*(4), https://doi.org/10.1016/j.cub.2014.11.068.

Meara, P. (2005). *Llama, Language Aptitude Tests: The Manual*.

Mikropoulos, T. A., & Natsis, A. (2011). Educational virtual environments: A ten-year review of empirical research (1999–2009). *Computers & Education*, *56*(3), 769–780. https://doi.org/10.1016/j.compedu.2010.10.020.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*(2), 81–97. https://doi.org/10.1037/h0043158.

Mimura, M., Komatsu, S., Kato, M., Yashimasu, H., Wakamatsu, N., & Kashima, H. (1998). Memory for subject performed tasks in patients with korsakoff syndrome. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *34*(2), 297–303. https://doi.org/10.1016/S0010-9452(08)70757-3.

Paas, F., & Sweller, J. (2012). An evolutionary Upgrade of cognitive load theory: Using the Human Motor System and collaboration to support the learning of Complex Cognitive tasks. *Educational Psychology Review*, *24*(1), 27–45. https://doi.org/10.1007/s10648-011-9179-2.

Paivio, A. (1990). *Mental representations: A dual coding approach*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195066661.001.0001.

Paivio, A. (2007). *Mind and its evolution: A dual coding theoretical approach*. Lawrence Erlbaum Associates Inc.

Paivio, A., & Csapo, K. (1969). Concrete image and verbal memory codes. *Journal of Experimental Psychology*, *80*(2), 279–285.

Paivio, A., & Csapo, K. (1973). Picture superiority in free recall: Imagery or dual coding? *Cognitive Psychology*, *5*(2), 176–206. https://doi.org/10.1016/0010-0285(73)90032-7.

Paivio, A., & Desrocher, A. (1980). A dual-coding approach to bilingual memory. *Canadian Journal of Psychology/Revue Canadienne De Psychologie*, *34*(4), 388–399.

Palmer, S. D., & Havelka, J. (2010). Age of acquisition effects in vocabulary learning. *Acta Psychologica*, *135*(3), 310–315. https://doi.org/10.1016/j.actpsy.2010.08.002.

Ping, R., & Goldin-Meadow, S. (2010). Gesturing saves Cognitive resources when talking about nonpresent objects. *Cognitive Science*, *34*(4), 602–619. https://doi.org/10.1111/j.1551-6709.2010.01102.x.

Poschner, J. (2018). *Vocabulary Acquisition Strategies & Language Aptitude BT - Exploring Language Aptitude: Views from Psychology, the Language Sciences, and Cognitive Neuroscience* (S. M. Reiterer (Ed.); pp. 245–259). Springer International Publishing. https://doi.org/10.1007/978-3-319-91917-1_13.

Pulvermüller, F. (1999). Words in the brain's language. *The Behavioral and Brain Sciences*, *22*(2), 253–279; discussion 280–336.

Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews Neuroscience*, *6*(7), 576–582. https://doi.org/10.1038/nrn1706.

Pulvermüller, F. (2018). Neurobiological mechanisms for semantic feature extraction and conceptual flexibility. *Topics in Cognitive Science*, *10*(3), 590–620. https://doi.org/10.1111/tops.12367.

Pulvermüller, F., Shtyrov, Y., & Ilmoniemi, R. (2005). Brain signatures of meaning Access in Action Word Recognition. *Journal of Cognitive Neuroscience*, *17*(6), 884–892. https://doi.org/10.1162/0898929054021111.

Rasouli, F., & Jafari, K. (2016). A deeper understanding of L2 vocabulary learning and teaching: A review study. *International Journal of Language and Linguistics*, *4*(1), 40. https://doi.org/10.11648/j.ijll.20160401.16.

Repetto, C., Colombo, B., Cipresso, P., & Riva, G. (2013). The effects of rTMS over the primary motor cortex: The link between action and language. *Neuropsychologia*, *51*(1), 8–13. https://doi.org/10.1016/j.neuropsychologia.2012.11.001.

Repetto, Claudia, Mathias, B., Weichselbaum, O., & Macedonia, M. (2021). Visual recognition of words learned with gestures induces motor resonance in the forearm muscles. *Scientific Reports, 11*(1), 17278. https://doi.org/10.1038/s41598-021-96792-9.

Repetto, C., Pedroli, E., & Macedonia, M. (2017). Enrichment effects of gestures and pictures on abstract words in a second language. *Frontiers in Psychology, 8*. https://doi.org/10.3389/fpsyg.2017.02136.

Repetto, C., Serino, S., Macedonia, M., & Riva, G. (2016). Virtual reality as an embodied tool to enhance episodic memory in elderly. *Frontiers in Psychology*, *7*(NOV). https://doi.org/10.3389/fpsyg.2016.01839.

Riesenberg, L. A., Leitzsch, J., & Little, B. W. (2009). Systematic Review of Handoff Mnemonics Literature. *American Journal of Medical Quality*, *24*(3), 196–204. https://doi.org/10.1177/1062860609332512.

Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences*, *20*(9), 676–688. https://doi.org/10.1016/j.tics.2016.07.002.

Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. *Trends in Neurosciences*, *21*(5), 188–194. https://doi.org/10.1016/S0166-2236(98)01260-0.

Robinson, P. (2001). Individual differences, cognitive abilities, aptitude complexes and learning conditions in second language acquisition. *Second Language Research, 17*(4), 368–392. https://doi.org/10.1177/026765830101700405.

Robinson, P. (2012). Aptitude in Second Language Acquisition. In *The Encyclopedia of Applied Linguistics*. https://doi.org/10.1002/9781405198431.wbeal0035.

Rogers, V., Meara, P., Barnett-Legh, T., Curry, C., & Davie, E. (2017). Examining the LLAMA aptitude tests. *Journal of the European Second Language Association*, *1*(1), 49–60. https://doi.org/10.22599/jesla.24.

Rueschemeyer, S. A., van Rooij, D., Lindemann, O., Willems, R. M., & Bekkering, H. (2010). The function of words: Distinct neural correlates for words denoting differently manipulable objects. *Journal of Cognitive Neuroscience*, *22*(8), 1844–1851. https://doi.org/10.1162/jocn.2009.21310.

Sadoski, M., & Paivio, A. (2012). *Imagery and Text A Dual Coding Theory of Reading and Writing*. https://doi.org/10.4324/9780203801932.

Schacter, D. L. (1987). Implicit memory: History and current status. In *Journal of Experimental Psychology: Learning, Memory, and Cognition* (Vol. 13, pp. 501–518). American Psychological Association. https://doi.org/10.1037/0278-7393.13.3.501.

Schmidt, M., Benzing, V., Wallman-Jones, A., Mavilidi, M. F., Lubans, D. R., & Paas, F. (2019). Embodied learning in the classroom: Effects on primary school children's attention and foreign language vocabulary learning. *Psychology of Sport and Exercise*, *43*, 45–54. https://doi.org/10.1016/j.psychsport.2018.12.017.

Seghier, M. L. (2012). The angular Gyrus: Multiple functions and multiple subdivisions. *The Neuroscientist*, *19*(1), 43–61. https://doi.org/10.1177/1073858412440596.

Skehan, P. (2002). No title. In P. Robinson (Ed.), *Individual Differences and Instructed Language Learning* (pp. 69–94). John Benjamins Publishing Company.

Skulmowski, A., & Rey, G. D. (2018). Embodied learning: Introducing a taxonomy based on bodily engagement and task integration. *Cognitive Research: Principles and Implications*, *3*(1), 6. https://doi.org/10.1186/s41235-018-0092-9.

Spada, N. (1997). Form-focussed instruction and second language acquisition: A review of classroom and laboratory research. *Language Teaching, 30*(2), 73–87. https://doi.org/10.1017/S0261444800012799.

Sweller, J. (1988). Cognitive load during problem solving: Effects on Learning. *Cognitive Science*, *12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4.

Tettamanti, M., Buccino, G., Saccuman, M. C., Gallese, V., Danna, M., Scifo, P., Fazio, F., Rizzolatti, G., Cappa, S. F., & Perani, D. (2005). Listening to Action-related sentences activates fronto-parietal motor circuits. *Journal of Cognitive Neuroscience*, *17*(2), 273–281. https://doi.org/10.1162/0898929053124965.

Tuena, C., Serino, S., Dutriaux, L., Riva, G., & Piolino, P. (2019). Virtual enactment effect on memory in Young and aged populations: A systematic review. *Journal of Clinical Medicine*, *8*(5), 620. https://doi.org/10.3390/jcm8050620.

Tulving, E., & Madigan, S. A. (1970). Memory and verbal learning. *Annual Review of Psychology*, *21*(1), 437–484. https://doi.org/10.1146/annurev.ps.21.020170.002253.

Turker, S., Reiterer, S. M., Schneider, P., & Seither-Preisler, A. (2019). Auditory cortex morphology predicts language learning potential in children and teenagers. *Frontiers in Neuroscience*, *13*(JUL). https://doi.org/10.3389/fnins.2019.00824.

Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition*, *92*(1), 231–270. https://doi.org/10.1016/j.cognition.2003.10.008.

Ullman, M. T., & Lovelett, J. T. (2016). Implications of the declarative/procedural model for improving second language learning: The role of memory enhancement techniques. *Second Language Research*, *34*(1), 39–65. https://doi.org/10.1177/0267658316675195.

Wang, Y. F., Petrina, S., & Feng, F. (2017). VILLAGE—Virtual immersive Language Learning and Gaming Environment: Immersion and presence. *British Journal of Educational Technology*, *48*(2), 431–450. https://doi.org/10.1111/bjet.12388.

Whiting, E., Chenery, H. J., & Copland, D. A. (2011). Effect of aging on learning new names and descriptions for objects. *Aging Neuropsychology and Cognition*, *18*(5), 594–619. https://doi.org/10.1080/13825585.2011.598912.

Willems, R. M., & Hagoort, P. (2007). Neural evidence for the interplay between language, gesture, and action: A review. *Brain and Language*, *101*(3), 278–289. https://doi.org/10.1016/J.BANDL.2007.03.004.

Yamamoto, Y. (2014). Multidimensional vocabulary acquisition through deliberate vocabulary list learning. *System, 42*, 232–243. https://doi.org/10.1016/j.system.2013.12.005.

Zimmer, H. D. (1996). Routes to actions and their efficacy for remembering. *Memory (Hove, England)*, *4*(1), 59–78. https://doi.org/10.1080/741940663.

Zimmer, H. D., & Saathoff, J. (1997). The influence of enactment on short-term recognition. *Acta Psychologica*, *95*(1), 85–95. https://doi.org/10.1016/S0001-6918(96)00030-3.

Authors and Affiliations

Manuela Macedonia[1] · Brian Mathias[2] · Astrid E. Lehner[3] · Susanne M. Reiterer[4] · Claudia Repetto[5]

✉ Manuela Macedonia
manuela.macedonia@jku.at

[1] Department of Information Engineering, Johannes Kepler University, Linz, Austria

[2] School of Psychology, University of Aberdeen, Aberdeen, United Kingdom

[3] Department of English and American Studies, University of Vienna, Vienna, Austria

[4] Department of Linguistics, University of Vienna, Vienna, Austria

[5] Department of Psychology, Università Cattolica del Sacro Cuore, Milan, Italy