Original Article

QJEP

Quarterly Journal of Experimental Psychology
1–11

DOI: 10.1177/17470218221082654
qjep.sagepub.com
SAGE

Comparing mnemonic effects of iconic gestures and pictures on word memory

Iván Sánchez-Borges and Carlos J Álvarez

Abstract

Previous studies using intermodal semantic priming have found that gestures improve language memory. In the present study, we ask whether the inherent characteristics of representative gestures (iconic gestures) facilitate word memory, or it is simply the semantic content shared with the words. Two analogous experiments were carried out presenting iconic gestures, pictures, or null primes to target words (nouns and action verbs). In Experiment 1, participants performed a free word recall task. In Experiment 2, the task was one of recognition. The results showed that participants recalled (Experiment 1) an equivalent number of words preceded by gestures or pictures compared with words alone, with no prime. However, a significantly higher number of words were recognised (Experiment 2) when they were primed by iconic gestures compared with the other two conditions, an advantage also found in reaction times (RTs) and both effects being larger in verbs than in nouns. These findings are discussed regarding the differences between recall and recognition processes as well as the particular characteristics of representative gestures.

Keywords

Iconic gesture; picture; memory; words

Received: 1 March 2021; revised: 4 February 2022; accepted: 7 February 2022

In everyday life, gestures are an essential part of the communicative process, often accompanying the use of spoken language. In addition, an important distinction has been made between representative or pantomimic gestures and non-representative or non-pantomimic gestures (e.g., Bernardis et al., 2008; So et al., 2013; Wu & Coulson, 2005, 2007a, 2007b; Yap et al., 2011). Non-representative gestures are random hand movements that accompany speech and do not transmit any obvious semantic meaning (McNeill, 1992). By contrast, an example of a representative gesture would be to extend the thumb to the ear and the little finger to the mouth, representing the gesture of "speaking on a telephone." Depending on the context, it can be accompanied by the words "call me" (an iconic gesture [IG]) or the gesture alone (pantomime) according to McNeill (2005) and following the work of Kendon (1980, 1982); However, most studies use the label IG broadly, even when the gesture is presented completely alone without speech (e.g., Kelly et al., 2010; So et al., 2013). The main objective of our study is to analyse whether representative gestures enhance word recall and/or recognition. For the reasons just given, and because we aim to study the influence of representative gestures on language memory (i.e., words), we will use the term IG, but admitting that it is debatable whether or not the term pantomime would be more appropriate.

In addition to the communicative effectiveness in both language production (e.g., Kita & Özyürek, 2003) and comprehension (e.g., Kelly et al., 2015), gestures can also have mnemonic benefits. In general, it has been found that the use of gestures is beneficial for language memory (Cohen & Otterbein, 1992; Feyereisen, 2006; Riseborough, 1981; So et al., 2012; Thompson, 1995), and when performed by the subject (subject-performed task [SPT]; Cutica & Bucciarelli, 2008; Cutica & Bucciarelli, 2013; Gimenes et al., 2013; Kartalkanat & Göksun, 2020). However, this possible benefit could depend on a gesture's meaning. Several studies have shown that pantomimic or representative gestures facilitate the recall of linguistic

Dpto. de Psicología Cognitiva, Social y Org. & IUNE, Universiy of La Laguna, San Cristóbal de La Laguna, Spain

Corresponding author:
Iván Sánchez-Borges, Dpto. de Psicología Cognitiva, Social y Org. & IUNE, Facultad de Psicología, University of La Laguna, Campus de Guajara s/n, La Laguna, 38071 San Cristóbal de La Laguna, Santa Cruz de Tenerife, Spain.
Email: ivansbr@hotmail.com

material compared with non-pantomimic or non-representative gestures (Cohen & Otterbein, 1992; Kelly et al., 1999; Thompson, 1995; Woodall & Folger, 1985), although non-representative ones could improve retention via attentional mechanisms such as those related to focusing attention or emphasis (So et al., 2012). The present research will be centred only on representative gestures.

Given the role of a speaker's gestures in language comprehension in real time, working memory is likely to play an important role. Working memory is known as the ability to store information for a limited period of time, while it is being processed (Baddeley, 2012). There are several models and theories that explain how information is stored in working memory. One of them is the classical multicomponent model proposed by Baddeley and Hitch (1974), a theoretical proposal which does not include or discuss gestures but that we consider relevant in this context. Working memory is critical for real-time processing, serving to temporarily maintain and store perceptual information, and allowing for adequate updating of long-term representations. In particular, working memory is thought to be composed of a central controller, the *central executive*, which is supported by two subsidiary systems: the *phonological loop*, capable of containing information based on verbal and written language, which, in turn, contains the *phonological store* that keeps the information for a few seconds, and the *articulatory rehearsal loop* that updates the stored elements and codes the new ones. The other system would be the *visuospatial sketchpad*, which is responsible for processing visual and spatial material. The two systems then form active stores that can combine information from sensory input and from the central executive. In its current form, the model postulates the existence of four components, including the visuospatial sketchpad, the phonological loop, the episodic buffer, and the central executive (Baddeley et al., 2011).

According to Baddeley and Hitch (1974; Baddeley et al., 2011), IGs, being a spatial or motor representation of a concrete action (McNeill, 1992), should be processed by the visuospatial sketchpad and not by the phonological loop, which oversees processing of verbal and written language. The same is applicable to pictures, another kind of non-verbal stimuli that can share meaning with words or sentences. There are several theoretical proposals that support this supposition. For example, the dual coding theory (DCT), although not specifically based on gestures, is a partial support for the interaction of verbal and non-verbal mental processes. Clark and Paivio (1991) also proposed that there are two psychological processes based on two independent but interconnected systems: the verbal system and the non-verbal system.

However, it has been shown that gestures, mainly non-representative ones, are not processed by the visuospatial sketchpad, as suggested previously (Baddeley & Hitch, 1974). According to Smyth et al. (1988), a spatial interference (e.g., tapping on four blocks arranged in a square) does not affect gesture retention, and only affects spatial performance and vice versa (see also Smyth & Pendleton, 1989). This suggests that gestures are processed by a different system to the visuospatial sketchpad. Therefore a new component to process and retain non-representative gestures has been proposed in working memory, in the context of the Baddeley and Hitch's (1974), Baddeley (2000) multicomponent model: the "Gestural Loop" (Gimenes et al., 2013; but see Wu & Coulson, 2014). Thus, as two different memory systems could oversee gestures and pictures, a comparison between them is interesting in itself. We will return to this issue later.

Another factor to take into account regarding the capacity of information retention in the working memory is the additive effect (Paivio, 1975). According to the DCT, when the information transmitted is redundant both verbally and non-verbally, it will leave a stronger imprint on memory (Clark & Paivio, 1991). It appears that information presented in a multimodal form improves information retention (Moreno and Mayer, 1999a, 1999b). For example, according to Mayer (1997), participants learn better (retain more information) when information is presented to them in images and words versus only words. In this sense, an IG can complement verbal information and leave a stronger trace of information in the memory, facilitating the recall of linguistic material such as sentences (Cohen & Otterbein, 1992; Cohen & Stewart, 1982; Feyereisen, 2006; Ianì & Bucciarelli, 2017; Nilsson & Craik, 1990). For example, Cohen and Otterbein (1992) and Thompson (1995) have found that sentences presented in conjunction with pantomimes showed greater recall than sentences presented without gestures. Also, Feyereisen (2006) found that participants recalled and recognised more sentences primed by representative gestures than by non-representative ones, though only when the IGs were congruent with the sentence meaning.

In sum, research has shown that sentences accompanied by representative and congruent gestures improve verbal memory and recognition. However, not much research has been devoted to exploring whether IGs improve word memory. One exception was Riseborough (1981), who asked participants to watch videos in which the narrator recited a list of verbs accompanied by IGs or non-gestures. The results showed that participants remembered more verbs that were accompanied by gestures than those that were not. Hupp and Gingras (2016) also investigated whether the memory of two types of words (nouns and verbs) were affected differently by gestures. Their hypothesis was that matching nonsense words (new words), whether verbs or nouns (e.g., "I took the zek from the library"), with IGs would facilitate the learning of that word, compared with beat gestures or nonsense gestures. They also investigated whether the frequency of the words influenced the accuracy of the memory. They predicted that nouns would show better memory than verbs, as nouns are conceptually more basic than verbs (Gentner, 1982). They concluded that IGs improved word learning compared with all the other gestural conditions. In addition, they observed that "high-frequency words" (new words with more exposures to the word/gesture combination) were recalled more accurate than "low-frequency words," and there were no significant differences between nouns and verbs.

Although most research has found that when a sentence or a word is previously accompanied by a congruent IG, recall and recognition are facilitated, some questions remain unsolved. In the studies reviewed above, a critical question has not been addressed: whether IGs actually help memory of words because their implicit properties or it is simply a matter of sharing semantic content with the word. To answer this question, we will compare the mnemonic effects of IGs with another non-verbal stimuli that also share semantic content with the word: pictures. In addition, a baseline will be used to have a clearer comparison: the presentation of the word alone.

As previously mentioned, in addition to the articulatory loop, the existence of a gestural loop has also been proposed but only for non-representative gestures (Gimenes et al., 2013). Furthermore, according to Gimenes's proposal, pictures would be processed in the visuospatial sketchpad whereas gestures in the gestural loop. In our two experiments, both IGs and pictures (primes) are always related to words (targets) and with a close and comparable semantic relationship with them. If IGs are processed differently from pictures, it is expected that there will be differences in the number of recalled and recognised words primed by one or the other.

Both IGs and pictures are perceived visually and spatially. In fact, according to McNeill (1992), gestures can be compared with pictures in the sense that both offer the opportunity to encode global and holistic relations, which contrasts with the analytic, linearly segmentable properties of speech. Although it is also obvious that gestures and pictures can vary in several dimensions and characteristics. Pictures, for instance, can provide details that gestures cannot. On the contrary, IGs can vary in their degree of semantic overlap with words. However, IGs are likely to be more enriched with dynamic spatial information than pictures, as pictures are usually perceived in a static way. In addition, as mentioned above, gestures provide temporality. According to McNeill (1992, 2005), a gesture is composed of three stages: preparation, stroke, and retraction. For example, when the "grab" gesture occurs, the person raises the hand to chest height (preparation phase), then extends the arm to the front and closes the fingers (stroke phase), and finally relaxes the hand until it returns to its initial position (retraction phase). These three stages generate the movement itself, providing a wider perception of the meaning compared with a picture. With this in mind, we expect that IGs, due to their greater temporal, dynamic spatial information and motor properties will improve both memory and word recognition compared with pictures or to only words. Furthermore, unlike most previous studies (e.g., Cohen & Otterbein, 1992; Feyereisen, 2006; Hupp & Gingras, 2016; So et al., 2012), both gestures and pictures will be presented in isolation and visually, that is, without the accompaniment of speech or other linguistic material.

Another of our aims is to explore whether the possible facilitation of IGs could be different across grammatical modalities (nouns and verbs). As mentioned, most of the previous studies have used sentences as targets. Only in the studies by Riseborough (1981) and So et al. (2012) did they use words (specifically verbs) and, as far as we know, only the study by Hupp and Gingras (2016) was the type of word (verb vs. noun) manipulated. We consider that the possible effect of the functional role of gestures on the grammatical roles of nouns and verbs deserves further investigation (see, for example, Bernardis et al., 2008; Goldin-Meadow & Wagner, 2005). IGs are primarily functional when the semantic meaning they convey is related to temporal, motor, and spatial information (Driskell & Radtke, 2003, and Hostetter, 2011). Therefore, this type of information could be more relevant and more prominent in action verbs than in nouns, even though the nouns refer to real objects, both kinds of stimuli were used in the two experiments.

In addition, the perspective of embodied cognition has strong implications for learning and information retention. Our focus here is on memory of verbal material. We specifically argue that memory of words preceded by gestures will be better than preceded by pictures or by nothing, as gestures are moving and motor stimuli, more related and closer to actions. Moreover, gestures are known to activate areas of the sensorimotor cortex (Martuzzi et al., 2014). Furthermore, an internal representation of gestures appears to be similar to inner speech, as observing movements activates the same cortical areas as performing them (Wilson & Emmorey, 2006; Wilson & Knoblich, 2005). Thus, it is likely that a word involving movement (e.g., an action verb) if preceded by a congruent gesture will facilitate the formation of a stronger memory trace compared with the presentation of the word preceded by a congruent picture or nothing (just the word).

Although in the study by Hupp and Gingras (2016) there were no significant differences between verbs and nouns, it is possible that the participants were learning new words, and these words were part of a sentence produced visually and aurally, so they might stop paying attention to gestures and focus only on the auditory modality. In our study, the linguistic stimuli will be isolated words. Therefore, a greater advantage for IGs is expected for verbs than for nouns, in both recall and recognition.

To our knowledge, no memory study has deliberately controlled for participants actually looking at the gestures that act as primes of linguistic targets (Cohen & Otterbein, 1992; Feyereisen, 2006; So et al., 2012). Only in the study by Gimenes et al. (2013) did participants have to reproduce the gestures they had observed, so participants were obliged to observe them. It is possible that participants might stop paying attention to the prime (gesture or picture), so it would not be entirely clear whether the gesture or picture would act as a prime. Therefore, in our two experiments, we add explicit questions about the relationship between the IG or picture and word in 10% of the trials. However, we recognise that this procedure is far from perfect as all the gestures were related to their target words and the participants could easily learn or be aware of this fact.

In addition, in the study by So et al. (2012), the presentation of the gesture had a duration of 3,000 ms, and in Hupp and Gingras (2016), between 5,000 and 6,000 ms. In the two experiments, the presentation times of the IGs and images acting as primes are shortened with the objective of minimising the probability of strategically recoding the gestural primes into verbal labels, following the logic and procedures of previous works (see So et al., 2013; Yap et al., 2011). In the two experiments presented here, the presentation time of the IGs and images acting as primes are as short as possible: 1,000 ms.

Another of our objectives is to check whether gestures facilitate both recall and recognition of words similarly. It is known that the memory processes that act on recall and recognition are not exactly the same, therefore some differences in the pattern of results are possible. Recognition is much closer to the daily situations where IGs interact with language. It can be achieved by means of two independent processes: the familiarity of the word and the memory itself (Mandler, 1980), whereas recall depends basically on conscious memory (Jacoby et al., 1993). In addition, recognition depends mainly on the specific processing of the element (Hunt & Einstein, 1981) and can be facilitated by performing the gesture, drawing the receiver's attention to certain words. Thus, a bigger facilitation of IGs on recognition compared with free recall could be expected.

In short, the main objective of this research is to examine whether representational gestures, specifically IGs acting as primes, influence the memory of linguistic units, that is, target words. More concretely, in free recall (Experiment 1) and recognition (Experiment 2). If the predicted improvement in the memory processes of words is due to intrinsic, differential, and specific characteristic of IGs, the expected facilitation should be greater in comparison with a no prime condition or with another kind of prime stimulus that also shares comparable or even bigger semantic relationship with the word: pictures. In addition, we test whether the effects are independent of grammatical modality: verbs and nouns. The possible benefit of IGs on recall and recognition rates and RTs should occur and be larger for action verbs than for nouns, as verbs more directly represent visuospatial, temporal, motor, and spatial features, a kind of information which is also more characteristic and salient in gestures.

Stimuli selection and normative representativeness study

To select the experimental stimuli, a preliminary normative study of both IGs and pictures was conducted prior to the experiments, as there was no study of this type in Spanish, mainly for IGs. After this selection, another study was carried out with the objective of comparing the representativeness or semantic relationships between IGs and pictures with their respective words.

A total of 80 IGs (40 for action verbs and 40 for nouns of objects) were recorded as single video clips. Each clip displayed the upper half of the body of an adult male actor performing a gesture of an action (verb) or an object (noun) with both hands and arms, and with a controlled duration (between 800 and 1,200 ms). For example, moving both hands forward as a representative gesture for "pushing", or moving both hands from the centre of the head to the ears as a gesture for "headphones." All videos were presented to a panel of judges (10 students who did not participate in the experiments), and they were asked to write a word (verb or noun) to describe the meaning of each of the gestures. The gestures were presented without audio. As speech was not available, the participants interpreted the meaning of the gestures according to their physical forms and movements. Only those pairs of gesture–word with an agreement above 90% by the judges were chosen. There were 48 words, 24 action verbs in the infinitive form (such as "push," "eat" or "drink") and 24 singular nouns ("headphone," "ring," "gorilla," etc.), and their corresponding IGs finally selected (see Supplementary Material).

For picture stimuli, 48 pictures were extracted from the ARASAAC (Aragonese Centre for Augmentative and Alternative Communication) database (Cabello & Bertola, 2015; Viera Delgado & Roldán Coya, 2018). The pictures were colour pictograms representing the same nouns and verbs already included in the IG selection. Each picture was presented in the centre of a monitor with a white background. For instance, a picture of "headphones" for that noun. Even when they were selected according to the words from the database, pictures were also presented to another panel of 10 judges, who were asked to write a word (verb or noun) to describe the meaning of each of the pictures. It was confirmed that agreement was above 90% as in the case of gestures.

The library of stimuli is available for non-commercial purposes through RIULL (Institutional Repository): videos (https://riull.ull.es/xmlui/handle/915/23607), pictures (https://riull.ull.es/xmlui/handle/915/25847).

After the previous selection, a quantitative study was carried out, comparing IGs and pictures in their relationship or representativeness with their respective words.

A sample of 106 undergraduate students of Psychology were randomly divided into two groups. There were 58 participants (46 females, 12 males; mean age = 20.8 years, range = 18–43) assigned to the IG-group and had to rate only a list with the 48 IGs and 49 participants (40 females, 9 males; mean age = 19.9 years, range = 17–32) were part of the Pic-group and only rated the 48 pictures.

Each group was asked to judge and rate the representativeness or relationship of each IG (IG-group) or picture (Pic-group) with the corresponding word. Both gestures and pictures were randomly presented on a computer screen one by one. In each trial, a video clip with the gesture (or the picture) was presented in the upper part of the screen, followed by the word (a noun or an action verb) and a 7-point Likert-type scale (1 = *not representative at all*, 7 = *very representative*). When the participant assigned a rating, the next trial appeared.

On average, the representativeness rating was 5.9 ($SD = .66$) for IG trials and 6.3 ($SD = .55$) for pictures. The unpaired *t*-test by participants and the paired *t*-test by items (with the scores for the pictures and for the IGs paired by their respective words) revealed that this difference was statistically reliable, t_1 (94) = 3.49, $p < .001$. t_2 (47) = 3.55, $p < .001$. In addition, *t*-tests for verbs and for nouns separately showed that the reliability came from nouns: 5.7 for IG and 6.6 for pictures, t_1 (46) = 6.24, $p < .001$. t_2 (23) = 6.35, $p < .001$. In the case of verbs, the difference between IGs (6.03) and pictures (6.09) was not significant, t_1 (46) < 1; t_2 (23) < 1.

Results showed that both kinds of stimuli were rated as highly representative of the words, with no significant difference in the case of verbs but with an advantage for pictures in the case of nouns. In addition, in the case of nouns, both IGs and pictures were highly representative of the respective words. Indeed, any superiority of the picture ratings would, in any case, go against our hypothesis.

Experiment 1

In most previous research, the mnemonic effect of representational gesture accompanied by sentences has been studied (Cohen & Otterbein, 1992; Cohen & Stewart, 1982; Feyereisen, 2006; Ianì & Bucciarelli, 2017; Nilsson & Craik, 1990). However, not much research has been devoted to exploring whether representational gestures facilitate memory processes in words. To our knowledge, only a few studies have used verbs (Riseborough, 1981; So et al., 2012) or verbs and nouns (Hupp & Gingras, 2016) in word recall tasks. However, none of these studies have compared the mnemonic effect of IGs with another stimulus that also shares semantic content with the two types of words or grammatical categories.

In this experiment, the free recall of words (nouns and verbs) preceded by IGs, pictures, or presented alone are compared. As previously stated, a superiority of IG over the other two prime conditions is expected. In addition, we consider that this advantage should be bigger for action verbs than for nouns.

Method

Participants. A total of 31 undergraduate students of Psychology and Speech Therapy (3 men and 28 women) from the University La Laguna, with Spanish as their first language, and with no history of neurological problems participated in the experiment to fulfil a course credit requirement. They were between 18 and 23 years old (*M*: 19.06 years).

Materials and design. The 48 words (24 verbs and 24 nouns) were presented preceded by either an always-congruent video clip (the action or gesture of pushing followed by the verb "push") or a congruent picture (the picture of a "square" followed by the word "square") or in isolation (no prime). Three counterbalancing sub-lists were generated (according to a Latin square), so that when a sub-list was presented in relation to the video clip in one condition, they were not presented in the other two conditions (picture and nothing), and the same for nouns. All words were preceded by videos or pictures, or nothing. No video or picture was repeated in the same list. Each participant received only one list. The presentation of each condition was random within each list.

Thus, the design was a factorial 3 × 2 within-subject design, with the factors being Type of Prime (video, picture, and no prime) and Type of Word (verb and noun). The number of recalled words (accuracy) was the dependent variable.

Procedure. The experiment was run on a computer using E-Prime 3.0 software (Kim et al., 2019). Each test session was conducted in a quiet room, free from noise. It began with precise instructions. Participants were informed that the experiment consisted of three parts. In the first part, they would see a series of stimuli, which could appear in three ways: (1) a picture followed by a word, (2) a video clip followed by a word (the video always showed IGs) or (3) just a word. They were asked to pay attention to the words for a later memory task. The second part consisted of performing a mathematical calculation task to avoid the retroactive interference of words (Baddeley, 1997; Loftus, 1977; So et al., 2012). The third and final part consisted of the recall phase. Participants were asked to remember as many words as possible from the first phase, writing them on a sheet of paper. There was no restriction on the time or order of the words to be recalled.

Table 1. Recalled words (in percentages) as a function of Type of Prime (video, picture, and null prime) and Type of Word (verbs vs. nouns) in Experiment 1.

	Verbs	Nouns
Video	43	49
Picture	40	42
Null prime	23	31

Each trial began with a fixation point (a cross) for 300 ms, presented in the centre of a 17-in. colour monitor, with a black background. Next, a blank screen for 200 ms was followed by the video clip in the centre of the screen (640 × 480 pixels) or by the picture (500 × 500 pixels), both for 1,000 ms. In the condition of null prime, the word was presented 500 ms after the fixation point. There was an interval of 500 ms before the start of the next trial.

To ensure that the videos and pictures were actually processed, in 10% of the trials, after the target presentation, the following question appeared on the screen: "IS THE VIDEO RELATED TO THE WORD?" or "IS THE PICTURE RELATED TO THE WORD?," depending on the Type of Prime, to which the participants had to answer "yes" or "no" using two labelled answer keys. The question was posed for a maximum of 5,000 ms, and the participants were also informed about this part during the instructions.

Results and discussion

The answers to the question about the video–word relationship showed that the participants had processed the videos: the rate of correct answers was above 93%.

The number of correctly recalled word (see Table 1) were analysed using linear mixed models, which simultaneously consider variability for participants and for items (Baayen et al., 2008; Bates, 2005). More concretely, this was performed by making use of the logit function for binomial data. Statistical software R was used with the lme4 package (Bates et al., 2014), and more specifically, the ULLRToolbox (Hernández-Cabrera, 2011). After checking that the sub-list of factors was not significant (required by the Latin square), the factors Type of Prime (video, picture, and null prime) and Type of Word (nouns vs. verbs) were entered as within-participant factors. The model was estimated according to Barr et al. (2013) with all repeated measures factors as fixed and random slopes across participants.

The priming effect was significant, χ^2 (2) = 15.71, $p < .001$. Post hoc analyses with Hochberg's adjustment showed that words primed by gestures were recalled better (46%) than words alone (27%), $z = 5.8$ $p < .001$. There was also an advantage of recalled words preceded by pictures (41%) compared with words alone or with null prime, $z = 5.03$ $p < .001$. However, the difference between gestures and pictures acting as primes was not significant, $z = 1.6$ $p > .05$.

Neither the Type of Word, χ^2 (1) = 0.28 $p > .05$ nor the interaction, χ^2 (2) = 2.13 $p > .05$ yielded significance.

According to previous research (e.g., Baddeley et al., 2011), all visual information is processed by the visuospatial sketchpad. The question we ask ourselves is whether this visuospatial information (different for IGs and pictures) really helps to store the information in the working memory or is simply sharing semantic content with the word. According to the present results, both IGs and pictures highly representative of the words facilitate similarly the recall of the semantically related words, regardless of the grammatical class (noun and verb), in comparison with words presented alone. It is true that in the previous normative study, an advantage in representativeness was found for pictures in comparison with IGs for nouns. We will return to this issue in the section "General discussion." Overall, it seems that the integration between verbal (word) and non-verbal information (IGs and picture) improves the memory of words. According to the DCT (Clark & Paivio, 1991), when verbal and related non-verbal information is coded, learning is more successful. For example, the presentation of an IG video related to the word can supplement verbal information, as opposed to presenting the word alone, thus leaving a stronger imprint on memory. Our results support this notion.

Furthermore, it seems that this facilitation of recall is independent of grammatical modality (e.g., Hupp & Gingras, 2016; Kelly et al., 2010; McNeill, 1992), as there was no difference in the facilitation of recall between nouns (objects) and verbs (actions) preceded by IGs or pictures. Therefore, although the semantic meaning of IGs included both temporal and spatial information (Driskell & Radtke, 2003), we have found no difference between grammatical class. The main effect of grammatical category is difficult to interpret and does not make sense theoretically, as psycholinguistic variables in nouns and verbs were not matched, as this was not our objective.

Experiment 2

In most previous research, only the mnemonic effects of gestures on free recall have been proven (Ianì & Bucciarelli, 2017, 2018; Riseborough, 1981; So et al., 2012). As recognition and recall are different processes, we consider that it is also relevant to check this issue. Only a few studies have used recognition tasks, either by recognising sentences (Feyereisen, 2006) or by recognising gestures (Cohen & Otterbein, 1992). The recognition task in the present study compared, again, the influence of representative gestures (IGs), pictures and only words on the individual's ability to discriminate new words from previously processed ones. In addition to the number of correctly recognised words, the RTs to words were analysed.

Method

Participants. A total of 30 undergraduate students of Psychology and/or Speech Therapy (5 men and 25 women), from the University La Laguna, with Spanish as their first language and with no history of neurological problems, took part in the experiment for study credits. The age range was between 18 and 24 years, with a mean of 19.1. None of them participated in Experiment 1.

Materials and design. Both the primes and the targets were the same as in Experiment 1.

Thus, the design was again a 3 × 2 factorial within-subject design, with the Type of Prime (video, picture, and null prime) and Type of Word (with two levels: verb and noun). In this experiment, both the response latencies or RTs for the words, and the number of correctly recognised words were the dependent variables.

Procedure. The whole procedure was identical to that in the previous experiment, but in this case, they were asked to pay attention to the words for a recognition task. In addition, RTs were also recorded. The word (target) was presented for a maximum of 3,000 ms or until the participant responded. Therefore, in the third and last part of the experiment, they were asked to press the YES key, as quickly as possible and without making mistakes if they had already seen the word in the first phase or press NO if they had not. In the recognition task, 48 new filler stimuli with the same characteristics of the experimental ones (24 nouns and 24 action verbs) were randomly presented mixed with the original stimuli.

As in the previous experiment, to ensure that the videos were processed, in approximately 10% of the trials, after the target presentation, the following question appeared on the screen: "IS THE VIDEO RELATED TO THE WORD?" or "IS THE PICTURE RELATED TO THE WORD?," depending on the Type of Prime, to which the participants had to answer "yes" or "no" using the two answer keys. The question was posed for a maximum of 5,000 ms.

Results and discussion

First, the answers to the question about the video–word relationship showed that the participants processed the videos with the rate of correct answers being above 95%.

The percentage of correctly recognised words and the RTs to words correctly recognised (see Table 2) were again analysed using mixed effects modelling. The factors were Type of Prime (video, picture, and null prime) and Type of Word (nouns vs. verbs). As in Experiment 1, for the number of recognised words, a logit function for binomial data was used, and the same model was estimated following Barr et al. (2013) with all factors of repeated measurements as fixed and random slopes across participants.

Analysis of mean percentage of recognised words showed a main effect of the Type of Prime χ^2 (2) = 36.27 $p < .001$. Post hoc analysis with Hochberg's adjustment showed that words preceded by gestures (87%) were better recognised than words alone (61%) $z = 8.1$ $p < .001$. The difference between words preceded by pictures (81%) compared with words alone was also significant, $z = 5.8$ $p < .001$. Notably, the difference between gestures and pictures acting as primes was significant, with a higher recognition rate for gesture than for picture primes, $z = 2.14$ $p < .05$. The Type of Word effect was also significant χ^2 (1) = 7, 94 $p < .01$: nouns were recognised in a higher proportion (81%) than verbs (71%).

Table 2. Mean values and standard deviations (in ms) for RTs and number of recognised words (in percentages, %) as a function of Type of Prime (video, picture and null prime) and Type of Word (verbs vs. nouns) in Experiment 2.

	Verbs		Nouns	
	Mean RT (*SD*)	%	Mean RT (*SD*)	%
Video	933 (378.4)	85	812 (277.6)	90
Picture	1,013 (499.1)	75	871 (389.4)	88
Null prime	1,095 (553.7)	55	895 (461.5)	67

RT: reaction time.

For the RTs, errors, and response latencies faster than 200 ms were excluded from the RTs analyses. Response latencies more than 2.5 *SD*s above or below each participant were also excluded from the analyses (3, 53% of the data in total).

The analysis, using the Satterthwaite approximation for degrees of freedom, showed a significant effect of the Type of Word, F (1, 42) = 21.32, $p < .001$: nouns were recognised faster (M = 856, SD = 375.6) than verbs (M = 1,002, SD = 474.8). The Type of Prime F (2, 70) = 6.82, $p < .01$ was also reliable. Post hoc analysis with Hochberg's adjustment showed that words preceded by gestures were recognised faster (M = 871.37 ms, SD = 335.56 ms) than words alone (M = 985.24 ms, SD = 514.2 ms), t (36) = 3.6 $p < .01$. There was also a difference between recognised words preceded by pictures (M = 936.2 ms, SD = 448.34 ms) compared with words alone or with a null prime, t (78) = 2.12 $p < .05$. Again, the difference between gestures and pictures acting as primes was significant, t (78) = 2.3 $p < .05$, with an advantage for gestures.

In general, the results about the Type of Prime were closer to our expectations than those of Experiment 1. IGs facilitated the processing of words in comparison with pictures or null priming, and for both measures: recognition accuracy and RTs for correctly recognised words. These effects were larger and clearer for verbs than for nouns.

General discussion

The main aim of our study is to examine whether IGs, due to their specific properties, have a mnemonic effect on word memory, facilitating word recall and recognition, or whether it is just the semantic content shared with the word that helps with linguistic information retention. To achieve this objective, we compared the possible mnemonic effects of IGs with another kind of visual stimuli that also shares meaning with words: pictures. The semantic relationship or representativeness between IGs and pictures with their corresponding words was previously tested and was quite high (with a significant advantage for pictures). In addition, the grammatical modality of the words (verbs and nouns) was manipulated. Thus, a cross-modal priming paradigm was used including words as targets. Pictures and IGs acted as primes and were compared with the presentation of the word alone, a baseline or control condition. The task was a classic one of free recall in Experiment 1 and a recognition task in Experiment 2. We expected that IGs would improve both recall and recognition of words compared with pictures (or to the words in isolation, with no primes) and that this effect would be larger in recognition than in recall and larger in verbs than in nouns. However, in Experiment 1 both pictures and IGs behaved similarly acting as primes facilitating recall compared with the word alone (null prime). By contrast, our prediction was confirmed in Experiment 2, both for RTs and for the numbers of correctly recognised words. There was a stronger facilitation effect of IGs compared with the other two conditions in the recognition of words, being the effect more robust for verbs than for nouns.

The plausible influence of gestures on the memory of verbal material has been widely studied. However, most research has used sentences as targets (e.g., Cohen & Otterbein, 1992; Feyereisen, 2006; Kelly et al., 1999; Thompson, 1995). Only a few studies have been restricted to words; a shorter and single unit with meaning (Hupp & Gingras, 2016; Riseborough, 1981; So et al., 2012). This second option was chosen for our two experiments because the expected influence of gestures can be studied more directly and precisely, leaving aside other variables that could influence memory processes (different words with diverse meanings and varied grammatical categories, context, etc.). Moreover, reading a sentence requires serial processing to subsequently recall each lexical unit and consequently, exceeds the capacity of the phonological loop and is therefore processed by the episodic buffer (Baddeley, 2000). According to Baddeley et al. (2009), the buffer is multidimensional, that is, it gathers related information from different sources including working memory and long-term memory. The capacity of the buffer memory is assumed to be limited by the number of multidimensional fragments it can contain at any given time. In sum, the memory process of a sentence preceded by a gesture or picture may not be as directly related to their semantic relationship as single words, due to the number of processes involved.

As previously mentioned, our main research question is: Do IG properties facilitate the memory of linguistic material, or is it simply the semantic content shared with the word? Not many studies have compared the mnemonic effect of IGs with another stimulus that also shares meaning with the word. In our study, we used pictures that, like IGs, clearly represented the target words. This allowed us to test whether it is the concrete visual and spatial information transmitted by gestures that facilitate memory of linguistic units. Alternatively, it could be just a question of a visual stimuli providing redundant, multimodal, complementary, or additional information to the words to be remembered, leading to stronger and better imprint on memory (Johnson et al., 1996).

The results of Experiment 1 (free recall), with no significant differences between gesture and picture conditions (although both facilitated recall in comparison with the null prime condition), suggest that the two prime stimuli benefit word memory in a similar way. However, and interestingly, when the task was recognition (Experiment 2), words primed by IGs were recognised faster and more accurately than when primed by pictures (or by null primes), being these effects stronger in verbs than in nouns. It is evident that both processes, recall and recognition, are different in several ways. Recognition is an activity much closer to language comprehension processes (i.e., discourse comprehension or reading), where new and stored information need to be continually integrated. In addition, recognition depends strongly on item-specific processing (Feyereisen, 2006; Hunt & Einstein, 1981). These differences between recall and recognition are possibly underlying our results, in which IGs produced a greater facilitation in word recognition than pictures, a difference not observed in free recall. In fact, previous studies have observed a similar pattern of results using pantomimic (i.e., representative or IGs) in comparison with non-pantomimic gestures, the first ones producing higher memory scores of sentences when the task was a different one to free recall (e.g., Cohen & Otterbein, 1992).

As tested in the preliminary selection and normative studies, gestures and pictures were semantically close to the target words, although the representativeness of the pictures was significantly higher than in the case of IGs only for nouns (the two kinds of visual stimuli were equally representative for verbs). This fact allows us to conclude that any benefit of IGs over pictures (as in Experiment 2) cannot be explained based on semantic proximity of both types of primes with regards to the words, or due to a difference in relationship with the linguistic material. In addition, if IGs as primes were more effective than pictures for nouns, the aforementioned superiority in the semantic relationship of pictures could have explained to some extent the null difference in the recall of nouns preceded either by pictures or by IGs in Experiment 1. However, there was also a similar recall score for both types of primes in the case of verbs, which strongly suggest that was not the case. By contrast, the superiority of IGs over pictures was observed in the two recognition measures of Experiment 2, and bigger for verbs than for nouns.

Moreover, the duration of the presentations of pictures and gestures cannot be producing these differences, as in the two experiments, the presentation time of both priming stimuli was matched (1,000 ms). In previous studies, the time interval of gestures was considerably longer (e.g., 5–6 s in Hupp & Gingras, 2016; 3 s in So et al., 2012). The duration of IGs in our study (1,000 ms) was kept as short as possible while still preserving a complete meaningful movement. The reason was to minimise or reduce the possible process of linguistically labelling the IGs, which in turn could lead to lexical or verbal primes rather than gestures acting as primes.

It is clear that IGs and pictures differ in a number of properties or dimensions. Although both are perceived and processed visually and both include spatial information, pictures are static stimuli, whereas IGs are enriched with dynamic, motor, and temporal information. In addition, gestures accompany language daily in communicative processes, and the empirical evidence of their positive influence over language processing (i.e., facilitation) is huge. On the contrary, in everyday life, a situation where a picture precedes a word or goes together with linguistic material (both interacting in a communicative process) is not as common as with a gesture. However, as previously remarked, the differences between pictures and gestures are not negligible. For instance, in certain circumstances pictures can make available to the system details or extra information that the IGs cannot. Thus, taking into account the present results, we consider that the beneficial and precise attribute (or attributes) of IGs that aid the recognition of words is a relevant issue that deserves to be further investigated in the future.

Furthermore, we predicted that this advantage of IGs would benefit more action verbs than nouns for several reasons. First, verbs are more difficult to learn than nouns because they are not as tangible. In fact, nouns are conceptually more basic than verbs (Gentner, 1982). For instance, it is easier to conceptualise the meaning of the noun "scissors" than the verb "to cut." But more importantly, the aforementioned traits of IGs based on dynamic information related to spatial, motor, and temporal features are also clearly more relevant and prominent in action verbs than in nouns.

Second, gestures are known to be proof or an indicator that the body is connected to verbal language and vice versa. Indeed, gestures have been considered as examples of embodied knowledge (Gibbs, 2005; Hostetter & Alibali, 2008; McNeill, 2005; Núñez, 2006). Studies on embodiment suggest that mental processes are mediated by body-based systems, such as motor, neural, and sensation and perception-based systems (Dreyfus, 1996; Glenberg, 2010). From the embodied cognition perspective, the influence or facilitation of IGs (i.e., motor stimuli made by the body and involving movement) on the memory of the same action verbs should be stronger than in the case of nouns. In our study, the recall and recognition of verbs and nouns was not differentially influenced by IGs. However, in Experiment 2, the advantage of the IG priming over the picture priming condition was stronger for verbs than for nouns, confirming largely our predictions. But the fact that IGs were superior as primes than pictures for both kinds of words (even when the last were of highest representativeness for nouns) is undoubtedly an interesting outcome, which would suggest a universal facilitation of gestures on different classes of words. But it is equally true that it is not clear how embodiment theories would explain the outcomes. However, elucidating this potential and theoretically relevant proposal is beyond the scope of the present work.

In sum, our results indicate that gestures, mainly iconic ones, facilitate the recognition of semantically related words compared with pictures or to the words in isolation. This facilitation is bigger in verbs than in nouns. However, when the task involves free recall, both pictures and IGs seem to act similarly, possibly because this memory task is farthest from the language comprehension processes, where item information, meaning, and specific characteristics need to be integrated with similar stored information. Our data suggest that the specific properties embedded in meaningful gestures (visual, spatial, dynamic, and temporal information) produce an additive effect improving word memory, as proposed some time ago by the DCT (e.g., Clark & Paivio, 1991).

Regarding working memory theoretical proposals, like the Baddeley et al. (2011) model, IGs would be processed and stored by the visuospatial sketchpad, as static images. However, it is equally true that pictures, although also sharing meaning with words, do not interact in everyday life with linguistic units as IGs do. On the contrary, a particular component has been put forward recently, known as the gestural loop (Gimenes et al., 2013). This specific working memory system would oversee processing and retaining gestures, although this mechanism would only work on non-pantomimic or non-representative gestures according to its proponents. Thus, the superiority of IGs over pictures observed in our data on word recognition could also reflect differences in two distinct systems, the gestural loop and the visuospatial sketchpad, only if the gesture-specialised mechanism processed representative gestures as well. A final limitation of our study deserves to be pointed out. The existence of differences in working memory among individuals has been repeatedly reported for a long time. For instance, recent research suggests that individual differences on visual working memory capacities are related to how much people benefit from gestures (Özer & Göksun, 2020; Wu & Coulson, 2014). Investigating this issue was not within our remit; although, we consider it as a main line for future experiments studying the influence of gestures on language memory.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was partially supported by Grant PID2020-114246GB-I00 from the Ministerio de Ciencia e Innovación (Spanish Government).

ORCID iD

Iván Sáchez Borges https://orcid.org/0000-0001-5359-1543

Supplemental material

The supplementary material is available at qjep.sagepub.com.

References

Baayen, R. H., Davidson, D. J., & y Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Baddeley, A. D. (1997). *Human memory: Theory and practice.* psychology press.

Baddeley, A. (2000). The episodic buffer: a new component of working memory?. *Trends in Cognitive Sciences, 4*(11), 417-423. https://doi.org/10.1016/S1364-6613(00)01538-2

Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology*, *63*, 1–29.

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.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *Psychology of learning and motivation* (Vol. 8, pp. 47–89). Academic Press.

Baddeley, A. D., Hitch, G. J., & Allen, R. J. (2009). Working memory and binding in sentence recall. *Journal of Memory and Language*, *61*(3), 438–456.

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

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*. https://doi.org/10.18637/jss.v067.i01

Bates, D. M. (2005). Fitting linear mixed models in R. *R News*, *5*, 27–30.

Bernardis, P., Salillas, E., & Caramelli, N. (2008). Behavioural and neuropsychological evidence of semantic interaction between iconic gestures and words. *Cognitive Neuropsychology*, *25*, 1114–1128.

Cabello, F., & Bertola, E. (2015). Características formales y transparencia de los símbolos pictográficos de ARASAAC. *Revista de Investigación En Logopedia*, *5*(1), 60–70.

Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review*, *3*(3), 149–210.

Cohen, R. L., & Otterbein, N. (1992). The mnemonic effect of speech gestures: Pantomimic and non-pantomimic gestures compared. *European Journal of Cognitive Psychology*, *4*(2), 113–139.

Cohen, R. L., & Stewart, M. (1982). How to avoid developmental effects in free recall. *Scandinavian Journal of Psychology*, *23*(1), 9–15.

Cutica, I., & Bucciarelli, M. (2008). The deep versus the shallow: Effects of co-speech gestures in learning from discourse. *Cognitive Science*, *32*(5), 921–935.

Cutica, I., & Bucciarelli, M. (2013). Cognitive change in learning from text: Gesturing enhances the construction of the text mental model. *Journal of Cognitive Psychology*, *25*(2), 201–209.

Dreyfus, H. L. (1996). The current relevance of Merleau-Ponty's phenomenology of embodiment. *The Electronic Journal of Analytic Philosophy*, *4*(4), 1–16.

Driskell, J. E., & Radtke, P. H. (2003). The effect of gesture on speech production and comprehension. *Human Factors*, *45*(3), 445–454.

Feyereisen, P. (2006). Further investigation on the mnemonic effect of gestures: Their meaning matters. *European Journal of Cognitive Psychology*, *18*(2), 185–205.

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. A. Kuczaj (Ed.), *Language development: Language, thought, and cul-ture* (pp. 301–334). Hillsdale, NJ: Erlbaum. ED219724.

Gibbs, R. W., Jr. (2005). *Embodiment and cognitive science*. Cambridge University Press.

Gimenes, G., Pennequin, V., & Sorel, O. (2013). Division of the articulatory loop according to sensory modality using double dissociation. *Journal of Cognitive Psychology*, *25*(7), 808–815.

Glenberg, A. M. (2010). Embodiment as a unifying perspective for psychology. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(4), 586–596.

Goldin-Meadow, S., & Wagner, S. M. (2005). How our hands help us learn. *Trends in Cognitive Sciences*, *9*(5), 234–241.

Hernández-Cabrera, J. A. (2011). *ULLR toolbox*. http://sites.google.com/site/toolbox

Hostetter, A. B. (2011). When do gestures communicate? A meta-analysis. *Psychological Bulletin*, *137*(2), 297–315.

Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin & Review*, *15*(3), 495–514.

Hunt, R. R., & Einstein, G. O. (1981). Relational and item-specific information in memory. *Journal of Verbal Learning and Verbal Behavior*, *20*(5), 497–514.

Hupp, J. M., & Gingras, M. C. (2016). The role of gesture meaningfulness in word learning. *Gesture*, *15*(3), 340–356.

Ianì, F., & Bucciarelli, M. (2017). Mechanisms underlying the beneficial effect of a speaker's gestures on the listener. *Journal of Memory and Language*, *96*, 110–121.

Ianì, F., & Bucciarelli, M. (2018). Relevance of the listener's motor system in recalling phrases enacted by the speaker. *Memory*, *26*(8), 1084–1092.

Jacoby, L. L., Toth, J. P., & Yonelinas, A. P. (1993). Separating conscious and unconscious influences of memory: Measuring recollection. *Journal of Experimental Psychology: General*, *122*(2), 139–154.

Kartalkanat, H., & Göksun, T. (2020). The effects of observing different gestures during storytelling on the recall of path and event information in 5-year-olds and adults. *Journal of Experimental Child Psychology*, *189*, Article 104725.

Kelly, S., Healey, M., Özyürek, A., & Holler, J. (2015). The processing of speech, gesture, and action during language comprehension. *Psychonomic Bulletin & Review*, *22*(2), 517–523.

Kelly, S. D., Barr, D. J., Church, R. B., & Lynch, K. (1999). Offering a hand to pragmatic understanding: The role of speech and gesture in comprehension and memory. *Journal of Memory and Language*, *40*, 577–592.

Kelly, S. D., Creigh, P., & Bartolotti, J. (2010). Integrating speech and iconic gestures in a Stroop-like task: Evidence for automatic processing. *Journal of Cognitive Neuroscience*, *22*(4), 683–694.

Kendon, A. (1980). Gesticulation and speech: Two aspects of the process of utterance. *The Relationship of Verbal and Nonverbal Communication*, *25*(1980), 207–227.

Kendon, A. (1982). The study of gestures: Some observation on its history. *Recherches Sémiotiques*, *2*, 45–62.

Kim, J., Gabriel, U., & Gygax, P. (2019). Testing the effectiveness of the Internet-based instrument PsyToolkit: A comparison between web-based (PsyToolkit) and lab-based (E-Prime 3.0) measurements of response choice and response time in a complex psycholinguistic task. *PLOS ONE*, *14*(9), Article e0221802.

Kita, S., & Öyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal? Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language*, *48*, 16–32.

Loftus, E. F. (1977). Shifting human color memory. *Memory & Cognition*, *5*(6), 696–699.

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, *87*(3), 252–271.

Martuzzi, R., van der Zwaag, W., Farthouat, J., Gruetter, R., & Blanke, O. (2014). Human finger somatotopy in areas 3b, 1, and 2: A 7T fMRI study using a natural stimulus. *Human Brain Mapping*, *35*(1), 213–226.

Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist*, *32*(1), 1–19.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago Press.

McNeill, D. (2005). *Gesture and thought*. University of Chicago Press.

Moreno, R., & Mayer, R. E. (1999a). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology*, *91*(2), 358–368.

Moreno, R., & Mayer, R. E. (1999b). Visual presentations in multimedia learning: Conditions that overload visual working memory. In *International conference on advances in visual information systems* (pp. 798–805). Springer.

Nilsson, L. G., & Craik, F. I. (1990). Additive and interactive effects in memory for subject-performed tasks. *European Journal of Cognitive Psychology*, *2*(4), 305–324.

Núñez, R. (2006). Do real numbers really move? Language, thought, and gesture: The embodied cognitive foundations of mathematics. In R. Hersh (Ed.), *18 unconventional essays on the nature of mathematics* (pp. 160–181). Springer.

Özer, D., & Göksun, T. (2020). Gesture use and processing: A review on individual differences in cognitive resources. *Frontiers in Psychology*, *11*, Article 573555.

Paivio, A. (1975). Coding distinctions and repetition effects in memory. *The Psychology of Learning and Motivation*, *9*, 179–214.

Johnson, C. J., Paivio, A., & Clark, J. M. (1996). Cognitive components of picture naming. *Psychological Bulletin, 120*(1), 113–139. https://doi.org/10.1037/0033-2909.120.1.113

Riseborough, M. G. (1981). Physiographic gestures as decoding facilitators: Three experiments exploring a neglected facet of communication. *Journal of Nonverbal Behavior*, *5*(3), 172–183.

Smyth, M. M., Pearson, N. A., & Pendleton, L. R. (1988). Movement and working memory: Patterns and positions in space. *The Quarterly Journal of Experimental Psychology*, *40*(3), 497–514.

Smyth, M. M., & Pendleton, L. R. (1989). Working memory for movements. *The Quarterly Journal of Experimental Psychology*, *41*(2), 235–250.

So, W. C., Low, A., Yap, D. F., Kheng, E., & Yap, M. (2013). Iconic gestures prime words: Comparison of priming effects when gestures are presented alone and when they are accompanying speech. *Frontiers in Psychology*, *4*, Article 779.

So, W. C., Sim Chen-Hui, C., & Low Wei-Shan, J. (2012). Mnemonic effect of iconic gesture and beat gesture in adults and children: Is meaning in gesture important for memory recall? *Language and Cognitive Processes*, *27*(5), 665–681.

Thompson, L. A. (1995). Encoding and memory for visible speech and gestures: A comparison between young and older adults. *Psychology and Aging*, *10*(2), 215–228.

Viera Delgado, B., & Roldán Coya, J. W. (2018). *Normative data on transparency of pictograms*. ARASAAC: estudio con población adulta.

Wilson, M., & Emmorey, K. (2006). No difference in short-term memory span between sign and speech. *Psychological Science*, *17*(12), 1093–1094.

Wilson, M., & Knoblich, G. (2005). The case for motor involvement in perceiving conspecifics. *Psychological Bulletin*, *131*(3), 460–473.

Woodall, W. G., & Folger, J. P. (1985). Nonverbal cue context and episodic memory: On the availability and endurance of nonverbal behaviors as retrieval cues. *Communications Monographs*, *52*(4), 319–333.

Wu, Y. C., & Coulson, S. (2005). Meaningful gestures: Electrophysiological indices of iconic gesture comprehension. *Psychophysiology*, *42*, 654–667.

Wu, Y. C., & Coulson, S. (2007a). How iconic gestures enhance communication: An ERP study. *Brain and Language*, *101*, 234–245.

Wu, Y. C., & Coulson, S. (2007b). Iconic gestures prime related concepts: An ERP study. *Psychonomic Bulletin and Review*, *14*(1), 57–63.

Wu, Y. C., & Coulson, S. (2014). Co-speech iconic gestures and visuo-spatial working memory. *Acta Psychological*, *153*, 39–50.

Yap, D. F., So, W. C., Melvin Yap, J. M., Tan, Y. Q., & Teoh, R. L. S. (2011). Iconic gestures prime words. *Cognitive Science*, *35*(1), 171–183.