UNIVERSITÀ DEGLI STUDI DI TORINO

Learning from text benefits from enactment

Ilaria Cutica[a], Francesco Ianì[b] & Monica Bucciarelli[b]

[a]Department of Health Sciences

Via A. di Rudinì, 8 – 20140, Milan, Italy.

e-mail: ilaria.cutica@unimi.it

[b]Department of Psychology and Center for Cognitive Science, University of Turin

Via Po, 14 –10123 Turin, Italy

e-mail: francesco.iani@unito.it

e-mail: monica.bucciarelli@unito.it

Corresponding author:
Ilaria Cutica
University of Milan
Department of Health Sciences
Via A. di Rudinì, 8
20140 Milan, Italy
e-mail: ilaria.cutica@unimi.it
tel: +39.02.50318938; fax:+39.02.91533899

Learning from text benefits from enactment

Abstract

Classical studies on enactment have highlighted the beneficial effect of gestures performed in the encoding phase on memory for words and sentences, for both adults and children. The present investigation focuses on the role of enactment for learning from scientific texts among primary school children. We assumed that enactment would favor the construction of a mental model of the text, and we verified the deriving predictions that gestures at the time of encoding would result in a greater number of correct recollections and discourse-based inferences at recall as compared to no gestures (Experiment 1) and in a bias to confound paraphrases of the original text with the text verbatim in a recognition test (Experiment 2). The predictions were confirmed; hence we argue in favor of a theoretical framework which accounts for the beneficial effect of enactment on memory for texts.

Introduction

Hand gestures are motor actions that often accompany speech, and are intertwined with the spoken content (e.g., Mc Neill, 1992; Kelly, Manning & Rodak, 2008). Studies on enactment have recognized a specific role of gestures in memory tasks, and purported that gestures enhance memory for speech. The term *enactment* refers to the finding that free recall of action phrases like "Break the toothpick" is improved when participants perform the action during the encoding phase (subject-performed task, SPT), compared to a situation where they read or hear the sentence (verbal task, VT) (Feyereisen, 2009). This effect has been consistent across numerous studies since the early 1980s (for reviews, see Engelkamp, 1998; Zimmer, 2001). A relevant finding is that the actual pattern of movements constituting SPTs is not critical in determining the recall level, as long as the patterns are appropriate to the accompanying speech (e.g., Cohen & Bryant, 1991; Noice & Noice, 2007). Noice and Noice (2007), for instance, detected the so-called non-literal enactment effect: even when the action performed is not literally congruent with the verbal material, but is related at a higher order level (e.g., at the action goal level), it results in action-enhanced memory for the verbal material.

The literature on gestures has also highlighted that the production of co-speech gestures by a learner is effective on memory when the gestures are produced in the encoding phase. Such *facilitating effect of co-speech gestures for the learner* may be viewed as analogous to the non-literal SPT effect. Producing gestures has been shown to play a key role in learning a variety of tasks (Broaders, Cook, Mitchell & Goldin-Meadow, 2007; Goldin-Meadow, Cook & Mitchell, 2009). For instance, Goldin-Meadow, Levine and colleagues (2012) invited the participants in their experiment to perform a mental rotation task and found that producing, rather than observing, gestures promotes learning as long as the gestures convey information that could help solve the task. Gestures have also been shown to play a key role in learning about math (Goldin-Meadow, Kim & Singer, 1999) and in conservation of quantity tasks (Ping & Goldin-Meadow, 2008).

The literature on the effect of SPTs on recognition and recollection memory for verbal material is mainly concerned with studies in which the participants deal with lists of words (e.g., Cohen, 1989; Frick-Horbury, 2002) or lists of phrases (e.g., Cohen, 1989; Feyereisen, 2006, 2009; Mangels & Heinberg, 2006; von Essen, 2005); few studies have dealt with memory for a discourse or a dialogue (Noice & Noice, 2001, 2007). It is beyond doubt that a text consists of a sequence of sentences, and that a discourse or a dialogue consists, more or less, of a sequence of spoken utterances. But texts, discourses and dialogues cannot be reduced to sentences and utterances, as their processing requires elements such as a context, cohesion, coherence, and rhetorical structure to be taken into account. All such elements impose meaning and structure on individual sentences (or utterances) that go well beyond the compositional meaning of sentences in isolation (see Graesser, Gernsbacher & Goldman, 2012). Furthermore, learning from a text or a discourse requires the ability to draw inferences about the information it contains, in order to generate links to establish inner text coherence (Albrecht & O'Brien, 1993), and also to draw inferences between all such information and the individual's previous knowledge. Given that comprehension and learning from verbal material also result in the ability to draw inferences, in our view, most of the studies on enactment, by failing to investigate and evaluate learning in terms of the ability to draw inferences, disregard a crucial aspect of learning from text. Therefore, our aim is to extend the investigation of the beneficial effects of enactment to learning from text.

2. The facilitating effect of gesture observation and gesture production on learning from text/discourse: A mental model account

The main assumption underlying our investigation is that gestures facilitate deep comprehension and learning from text or a discourse because they favor the construction of a text/discourse mental model. It is well known that, in comprehending a text or a discourse, people construct a mental representation on the basis of the semantic and pragmatic information contained in the text, together with their prior knowledge, and any inferences that are drawn; generally, such

mental representations do not contain surface information (the linguistic form of sentences). According to different theoretical frameworks, such representations are referred to as the "mental model" (Johnson-Laird, 1983; 2006) or "situation model" (van Dijk & Kintsch, 1983; an extension is the Construction Integration Model of Comprehension - Kintsch, 1998). For our purposes, we consider the two terms to be equivalent, disregarding their different theoretical roots (see also Kaup, Kelter & Habel, 1999).

The construction of a coherent mental model is tantamount to the successful comprehension at the text/discourse level (e.g., Glenberg, Kruley & Langston, 1994; Graesser, Millis & Zwaan, 1997; McNamara, Miller & Bransford, 1991; Zwaan, Magliano & Graesser, 1995; Zwaan & Radvansky, 1998), in that it integrates temporal, spatial, causal, motivational, person- and object-related information stated explicitly in the text. More in detail, in text/discourse comprehension people construct a model for each sentence, integrate such models also taking into account their prior knowledge, and consider what, if anything, follows (discourse-based inferences). Several studies investigating mental models in narrative comprehension have emphasized the spatial properties of situation models (Bower & Morrow, 1990; Glenberg, Meyer & Lindem, 1987; see also Jahn, 2004). In line with such view, we argue that co-speech gestures, which are spatial in nature, convey information that can be easily incorporated into the text/discourse mental model because mental models themselves are spatially organized (Knauff & Johnson-Laird, 2002); moreover, gestures are cast in the same non-discrete representational format as mental models. Previous studies have shown that co-speech gestures performed by the speaker facilitate the construction of an articulated mental model by the listener (Cutica & Bucciarelli, 2008, 2011; this also holds for oral deaf individuals trained to lip-read: Vendrame, Cutica & Bucciarelli, 2010). A better mental model results in a greater number of correct recollections and correct inferences drawn from the information explicitly contained in the discourse, and poorer retention of surface information (verbatim). These findings are consistent with the previous literature on mental models (see, e.g. Bransford, Barclay & Franks, 1972; Bransford & Franks, 1976; Johnson-Laird, 1983; Johnson-

Laird & Byrne, 1991; Johnson-Laird & Stevenson, 1970) claiming that mental models encode little or nothing of the linguistic form of the sentences on which they are based, and that, as a consequence, individuals tend to confuse inferable descriptions with the originals (e.g., Mani & Johnson-Laird, 1982).

As regards gesture production in learning from a text, the non-literal enactment effect (Noice & Noice, 2007) describes the beneficial effects of producing gestures relevant to the verbal material to be learnt. In our view, such enactment favors the construction of an articulated mental model of the text, thus improving memory for the content. A previous study (Cutica & Bucciarelli, 2013) showed that when adults, in the learning phase, produce gestures to enact the concepts contained in a text, at recall their memories reflect the construction of a more complete and integrated mental model, as compared to when they do not produce such gestures.

As learning from text starts to become relevant at school age, we extended our analysis to include primary school children. This extension is not trivial, since we know from the literature that the relationship between gestures and speech in children may differ from that in adults. For instance Alibali, Evans, Hostetter and colleagues (2009) found that children (aged 5 to 10 years) produce more non-redundant gesture–speech combinations than adults, both at the clause level and at the word level, suggesting that gesture–speech integration is not constant over the life span, but instead appears to change with development. With regard to the gestures that accompany oral narratives, some studies involving children aged 6–11 years (Kunene & Colletta, 2007; Colletta, 2009) found that only older children (9 years and over) use co-speech gestures in a similar way to adults (i.e., to represent the narrated events and characters' attitudes, to mark discourse cohesion and the pragmatic framing of the utterance). Colletta, Pellenq and Guidetti (2010) reported that spontaneous gesture production increases steadily with age: the adults in their study gestured slightly more than the 10-year-olds who, in turn, gestured more than the 6-year-olds. Such results support the claim that gesture production develops with age. Given the differences between adults' and children's use of gestures, our investigation aimed to verify whether gesture production has beneficial effects on

learning from scientific texts in children of school age, as is the case for young adults (see Cutica & Bucciarelli, 2013).

Several studies have shown that gestures may facilitate learning for children as well as for adults. Wagner Cook and Goldin-Meadow (2006), for example, found that fourth grade children (9-10 years old) who produced gestures during instruction on a math task were more likely to retain and generalize the knowledge they gained, than children who did not gesture. Moreover, Wagner Cook, Mitchell and Goldin-Meadow (2008) found that when third and fourth grade children were asked to instantiate a new concept in their hands, learning was more lasting than when they were asked to instantiate it in words alone. The authors concluded that gesturing seems to play a causal role in learning, by giving learners an alternative, embodied way of representing new ideas. Our experiments, as well as those in the relevant literature we have mentioned, involved experimenter-imposed gestures rather than gestures produced spontaneously. Spontaneous gestures may also be produced without conscious attention whereas gestures produced on demand focus the speaker's attention on both the speech whose content they should represent, and on the gestures themselves, which represent the text information in a non-discrete and spatial format. Such a focusing effect further supports the facilitating effect of gestures.

We conducted our experiment on 10-year-old children. This age group was chosen because the children had already acquired good reading skills and dealt with scientific texts at school and, at the same time, they still used gestures in a different way to adults. In Experiment 1 we tested the prediction that when children read a text while gesticulating, they would be more likely to recall correct information and draw discourse-based inferences than when they read it without gesticulating. In Experiment 2 we tested the prediction that when asked to gesticulate while reading a text, children would show a stronger bias to confound paraphrases with the textbase than when asked to keep their hands still. This prediction would appear to contradict findings in the enactment literature according to which enactment of single phrases enhances verbatim memory. But the contradiction is only apparent, and can be explained by taking into account the above mentioned

differences between sentence and text processing. In the gesture condition of our study, the children enacted concepts expressed by sentences within a compound text; therefore, to enact them, they also had to take into account the relevant concepts in previous sentences. In other terms, as they were forced to build a single mental model incorporating all the concepts expressed in the text, they consequently were more likely to lose track of the surface form of the single sentences.

Experiment 1: Gestures at the time of learning favor subsequent recall from text (recall-test)

The children in our experiment were asked to read and study two scientific texts, one while producing gestures to represent the information in the text and one while keeping their hands still. Then they were asked to recall as much information as they could. We predicted that when children produced gestures while reading a text (Gesture condition) they would retain more correct information and draw more discourse-based inferences than when they kept their hands still (No-Gesture condition). As regards inferences, we distinguished between discourse-based and elaborative inferences, as only the former are based on mental models. Discourse-based inferences make explicit that information which is originally implicit in the text; they may regard, for instance, the causal antecedent, the causal consequent, and the character's mental states with respect to the actions described (Graesser, Singer & Trabasso, 1994). They also establish the coherence between the part of the text that is being processed, and the previous one. Elaborative inferences (e.g., Singer, 1994) are instead a sort of enrichment of the original text; they do not depend upon the coherence of the text, nor do they enhance it. We made no predictions about the number of elaborative inferences, as these inferences may not be considered indices of a good mental model construction, because they are only arbitrary text enrichments. Furthermore, since in the gesture condition the children were invited to enact concepts that were interrelated insofar as part of a unique text, the positive effect of enactment on the literal recollection of single phrases should not apply in this condition. Hence, we predicted no increase in literal recollections in the Gesture condition as compared to the No-Gesture condition. Finally, we made no predictions about

erroneous recollections because the construction of a mental model, per se, does not guarantee that mistakes will not occur.

METHOD

Participants

Twenty-four children attending the 5th class of an Italian primary school (12 females and 12 males; mean age:10.9 years) participated in the study. Each child was assigned to both conditions (Gesture and No-Gesture conditions). All the children's parents had given their written consent prior to their participation in the study.

Materials

The experimental material comprised two short scientific texts: one about the circulatory system (Appendix A.1) and the other about the pulling force (Appendix B.1). Each text was 204 words in length. None of the children taking part in the experiment had already studied the topics of the two scientific texts in the classroom with their teachers.

Procedures

Each child encountered both scientific texts, one in the Gesture condition and the other in the No-Gesture condition. The occurrence of the two texts in each condition was counterbalanced over all participants. Further, half of the participants dealt with the Gesture condition first and half with the No-Gesture condition.

The experiment was run individually in a quiet room, with just the experimenter present. The experimenter introduced the experiment as a game as follows: "I'm going to ask you to play a game, an important game to figure out how we remember things that we read. This isn't an exam and therefore you won't get a mark, but the results will help us to understand how our mind works". Then, in the *Gesture condition* the children were instructed as follows: "Please read this text out loud and study it. While you are reading it, use your hands to help yourself. Try to represent what you are reading with your hands. Try to say the things you read with your hands". To make sure that the children had understood the instructions, before reading the text and by way of

example they were asked to use their hands to represent an apple on the table and a flying kite. Then they were asked to read the text twice, out loud. At the end of the second reading, they were asked to recall as much information as they could through the following instruction: "Now please tell me everything you can remember about the text you have just read, giving me as many details as possible". In the *No-Gesture condition* the children were instructed as follows: 'Please read this text out loud and study it, keeping your hands still and on the table'. The children were asked to read the text twice. At the end of the second reading, they were asked to recall as much information as they could. Each child's experimental session, comprising both the Gesture and the No-Gesture conditions, was video recorded.

Two independent judges coded the children's recollections. To this aim, each text was divided into 18 semantic units, corresponding to the main concepts that the children could recall. In particular, each concept (i.e. semantic unit) recalled by the participants was evaluated according to the following coding schema:

- *Literal recollection*: a semantic unit recollected in its literality.
- *Correct recollection*: a semantic unit recollected as a paraphrase.
- *Discourse-based inference*: a recollection in which the participant gave explicit information that was originally implicit in the semantic unit.
- *Elaborative inference*: the addition of plausible details to a semantic unit.
- *Erroneous recollection*: a recollection the meaning of which was inconsistent with the semantic unit.

To clarify the coding of the types of recollections consider, for instance, the following semantic unit: "It is due to the force of gravity that when we lose our balance, we fall to the ground"; according to the coding schema, the statement "When we lose our balance, we fall to the ground" is a Literal recollection. Consider now the following semantic unit: "The pulling force exerted by the Moon on the Earth is, instead, much smaller, because the mass of the Moon is much smaller than that of our planet". According to the coding schema, the statement "The Moon has a

weaker pulling force because its mass is smaller than that of our planet" is a Correct recollection; the statement "The Earth, that has a smaller mass, has a weaker pulling force" is an Erroneous recollection. The statement "The pulling force that the Earth exerts on the Moon is greater, because the Earth has a larger mass than the Moon" is a Discourse-based inference. Finally, the sentence "The Moon has a weaker pulling force and for this reason its atmosphere isn't compatible with human life" is an Elaborative inference. A partial recall of a semantic unit was scored by lenient criteria. Thus, for example, a semantic unit recollected as a paraphrase but with omission of details was coded as Correct recollection (see, e.g., the example of Correct recollection above).

In addition, two independent judges, who differed from those who coded the verbal recollections, coded the gestures produced by the children in the Gesture condition while reading the texts according to the following coding schema:

- *Representational gestures*: gestures that pictorially represent either concrete or abstract concepts (i.e., iconic and metaphoric gestures, respectively);
- *Deictic gestures*: indicative or pointing acts, commonly used to indicate people, objects, directions, and places, whether real, i.e., which exist in the space around the speaker, imaginary, or abstract (for example, things that have already been mentioned in the discourse);
- *Beats* or *motor gestures*: rhythmic or repetitive movements that are not related to the semantic content of the accompanying words or sentences, but are coordinated with the speech prosody and fall on stressed syllables;
- *Symbolic gestures*: conventional gestures whose meaning is culturally defined.

To clarify the coding of the types of gestures, consider, for instance, the following examples. Making a fist while reading the text "it (the heart) is about the size of a fist" is a representational gesture. Pointing to one's heart while reading the text "The heart is the motor of the circulatory system" is a deictic gesture. Simply moving the hand three times on the table while reading "red

blood cells, white blood cells, platelets" is a beat. Moving the index finger to the right and left while reading the word "not" in the sentence "which are not communicating" is a symbolic gesture.

Results

Two independent judges coded the participants' recollections individually; they reached a significant level of agreement on their first judgments for the overall group of participants in each experimental condition, calculated using Cohen's K (.91 < *K* < .95, *p* always <. 0001). For the final score the judges discussed each item on which they disagreed, until reaching a full agreement.

We checked the normality assumption using the Kolmogorov-Smirnov test. The frequencies of the different types of recall by the children in the experiment were, in actual fact, not normally distributed. The Kolmogorov-Smirnov test determined that the frequencies of types of recall in both the Gesture and No-Gesture conditions did significantly differ from the normal distribution (KS-test: df(24), *d* varied from .54 to .25, *p* varied from <.0001 to =.001). Statistical analyses were thus performed using nonparametric statistical tests.

The two texts were comparable in difficulty: we detected no difference in the quantity of types of recollections for the two text contents (Mann-Whitney test: *z* value varied from 1.59, to .03, *p* varied from .11 to .97). Hence, we pooled together the results for the two texts. Table 1 shows the mean scores for types of recollections in the two experimental conditions.

Insert Table 1 about here

In line with our predictions, there were more Correct recollections and Discourse-based inferences in the Gesture condition than in the No-Gesture condition (Wilcoxon test: *z*=2.48, *tied*[1] *p*=.007, and *z*=1.67, *tied p*=.048, respectively) albeit the *p* value for the Discourse-based inferences is close to .05, whereas Literal recollections occurred to the same extent in the two conditions (*z*=.28, *p*=.78). Also, there were fewer errors in the Gesture condition than in the No-Gesture

[1] "Tied p" is the value of p to be considered when the test is one tailed, that is when there is a directional hypothesis.

condition (z= 2.44, p=.015). Elaborative inferences occurred to the same extent in the two conditions (z=.45, p=.66).

Two more independent judges coded the gestures performed by the participants in the two readings of each text. They reached a significant level of agreement on their first judgments for the overall group of participants in the two experimental conditions, calculated using Cohen's K (K=.95, p<.0001). For the final score the judges discussed each item on which they disagreed, until reaching a full agreement. The participants in the Gesture condition performed a comparable number of gestures while studying the circulatory system and the pulling force texts (over the two readings, a mean of 31 and 33, respectively; Mann-Whitney test: z=.84, p=.40). Hence, we pooled the results for the two texts. They also performed a comparable number of gestures in the first reading (a mean of 33) and in the second reading (a mean of 32; Wilcoxon test: z=.43, p=.67). Hence, we pooled the results for the two readings.

The participants in the Gesture condition followed the instruction to accompany the reading of the semantic units in the text with gestures. Twelve of the 24 children produced gestures for all 18 semantic units in the text, seven children did so for 17 semantic units, four children for 16 semantic units, and one child for 12 semantic units. The children produced a considerable number of gestures (over the two readings, a mean of 54 representational gestures, four deictic gestures, six beat gestures, and one symbolic gesture). They often accompanied one semantic unit with several gestures, also of different kinds.

To perform an exploratory analysis to ascertain whether certain types of gestures favored learning from text more than others, we recoded the types of recollections as follows: Literal recollections; Proper recollections, comprising Correct and Discourse-based recollections, and Wrong recollections, comprising Elaborative inferences and Errors. Table 2 illustrates the mean percentages of types of recollections (Literal, Proper and Wrong) as a function of the type of gesture produced while reading. The percentages in each column do not add up to 100 because some children may have performed more than one type of gesture in correspondence with each semantic

unit or no gesture at all. The table should be read as follows. Consider for example Proper recollections as a function of representational gestures while reading; 51% of Proper recollections were accompanied by representational gestures while reading and the counterbalance of 51%, i.e. 49%, were not.

Insert Table 2 about here

As a general result, representational gestures were produced in correspondence with those semantic units that were more likely to be recalled as Correct recollections than as Literal or Wrong recollections (Wilcoxon test: z=3.88, p<.0001, and z=3.79, p<.0001, respectively). Moreover, semantic units for which children produced beats were more likely to be recalled as Proper recollections than as Literal recollections (Wilcoxon test: z=2.67, p=.008), and semantic units for which children produced deictic gestures were more likely to be recalled as Proper recollections than Wrong recollections (Wilcoxon test: z=2.20, p=.028). All the other comparisons yielded non-significant differences (Wilcoxon test: z value ranging from 0 to 1.73, and p value ranging from 1 to .083).

For explorative purposes, we also analyzed the types of gestures produced by children in the recall phase of the Gesture condition to see whether there was a correspondence between the gestures accompanying a semantic unit while reading and the gestures performed when recalling that semantic unit. The analysis concerned the performance of 21 children because 3 children did not produce any gesture at recall. Two independent judges analyzed the videos of the reading session and the recall session of each participant. In particular, they judged whether each gesture produced at recall in correspondence with a specific semantic unit was formerly performed by the participant in the reading phase in correspondence with the very same semantic unit, either in the first or the second reading. The two judges reached a significant level of agreement on their first judgments, calculated using Cohen's K (K=.71, p<.0001). For the final score the judges discussed each item on which they disagreed, until reaching a full agreement.

The results revealed that the 21 children who gestured at recall in the Gesture condition produced a mean of 5.6 gestures, and that 53% of such gestures differed from those produced while reading. A detailed analysis by single participant revealed that, at recall, nine children produced more gestures that were identical to those produced while reading than gestures that were different, eight children produced more gestures that were different from those produced while reading than gestures that were identical, whereas four out of 21 children produced about the same amount of gestures that were identical to and different from those produced at study. The only conclusion we can draw from these results is that at recall children did not tend to reinstate the gestures produced while reading.

Discussion

As predicted, we found that when children were invited to gesture to represent the concept they were learning, they produced more Correct recollections and more Discourse-based inferences at recall. In particular, children produced very few Discourse-based inferences, but all in the Gesture condition: five of the 24 children did so, whereas none of them drew Discourse-based inferences in the No-Gesture condition. In our view Correct recollections reveal that the individual has built an articulated mental model of the material to be learnt; the result for Discourse-based inferences is not as conclusive, but goes in the same direction. Furthermore, our results showed Literal recollections to be equally present at recall in the two conditions; this suggests that the facilitative effect we found on recall memory does not depend on the same process that sustains the enactment effect on memory for sentences. This may depend, as expected, on the fact that the material we used forced the integration of sentences into a unifying mental model, thus enhancing deep processing.

We also found that children who gesticulated while reading made fewer mistakes; this result was not predicted, but it reinforces our finding according to which the possibility of representing the concepts to be learnt with gestures results in a better quality of the mental model. Finally, we also checked the production of Elaborative inferences, even though these should not depend on the

construction of a mental model; we observed that children produced this kind of inference at the same rate in the two conditions.

The exploratory analysis of gestures performed in the encoding phase with respect both to the type of recollection and to the type of gestures performed at recall did not allow us to draw any conclusion.

Experiment 2: Gestures at the time of learning lead to confounding paraphrases with verbatim text (recognition test)

The children in the experiment were asked to read and study two scientific texts, one while representing the information in the text through gestures and one without gesticulating. After studying each text, they were presented with a series of sentences and, for each one, were asked to state whether it was the original sentence in the text. We predicted that when children produced gestures while reading a text (Gesture condition) they would be more likely to accept paraphrases of the original sentences as if they were the original sentences than when they did not (No-Gesture condition). We made no specific predictions on the ability to recognize literal sentences; according to our assumption on mental model construction, in the Gesture condition children should build a more articulated mental model at the expense of surface form than in the No-Gesture condition, and thus have more difficulty in recognizing the literal sentence. However, the findings of previous enactment studies have shown that, in memory for sentences, individuals who enact concepts are more able to recognize the literal form of sentences. Although our material did not consist of single sentences, we do not have enough elements to predict how such tendencies may interact in our task.

METHOD

Participants

Twenty-for children attending the fifth class of Italian primary school (10 females and 14 males; mean age: 10.6 years) participated in the study. Each child dealt with the Gesture and the No-Gesture conditions. All the children's parents gave their written consent prior to their participation in the study. None of the children had taken part in Experiment 1.

Materials and Procedure

The materials consisted of the scientific texts used in Experiment 1. For each text we chose six sentences and for each one we devised a triplet: Literally correct (the very same sentence presented in the text); Paraphrase (a sentence with the same meaning, but expressed using different words); Wrong content (a sentence inconsistent in meaning). We thus obtained 18 sentences, with six in each category: Literally correct, Paraphrase and Wrong content (see Appendix A.2 and B.2 for examples). Each triplet consisted of the same number of words. The experiment was run individually and in a quiet room with just the experimenter present, who introduced the experiment as a game. In both conditions, Gesture and No-Gesture, the children were asked to read each text twice, out loud. At the end of each reading the children were asked to read the 18 sentences one by one (presented in two different random orders over all participants) and to consider whether each of them was exactly the same as one of those in the text they had just read.

We coded 'Yes' responses to Literally correct sentences and 'No' responses to Paraphrase and Wrong content sentences as correct. Two independent judges also coded the gestures produced by the children while reading the texts in the Gesture condition; the judges followed the same coding schema as in Experiment 1.

Results

We checked the normality assumption using the Kolmogorov-Smirnov test. The frequencies of the different types of recognition by the children in the experiment were, in actual fact, not normally distributed. The Kolmogorov-Smirnov test determined that the frequencies of types of recognition in both the Gesture and the No-Gesture conditions did significantly differ from the

normal distribution (KS-test: df(24), *d* varied from .27 to .21, *p* varied from <.0001 to .008). Statistical analysis were thus performed using nonparametric statistical tests.

The two texts involved the same degree of difficulty; we detected no difference in the number of correct answers with the three types of sentences in the texts on the circulatory system and pulling force (Mann-Whitney test: *z* varied from .19 to 1.66, *p* varied from .85 to .10). Hence we pooled the results for the two texts. Table 3 illustrates the mean scores for the types of sentences accepted by the participants in the Gesture and No-Gesture conditions.

Insert Table 3 about here

As predicted, a comparison between the two conditions revealed that children were more likely to accept Paraphrases in the Gesture condition than in the No-Gesture condition (Wilcoxon test: z=2.14, tied p<.02), and equally likely to accept Literal sentences (z=.20, p=.84). The children were also equally likely to accept Wrong content sentences (z=.36, p=.72).

As regards accuracy, the children's performance with Paraphrases was better in the No-Gesture condition than in the Gesture condition (Wilcoxon test: z=2.14, p=.033), whereas their performance with Literal and Wrong sentences was comparable in the two conditions (Wilcoxon test: z=.20, p=.84, and z=.36, p=.72).

Two more independent judges coded the gestures performed by the participants in the two readings of each text. They reached a significant level of agreement on their first judgments for the overall group of participants in the two experimental conditions, calculated using Cohen's K (K=.93, p<.0001). For the final score the judges discussed each item on which they disagreed, until reaching a full agreement. In the Gesture condition, the children performed a comparable number of gestures while studying the circulatory system and the pulling force texts (over the two readings, a mean of 33 and 31, respectively; Mann-Whitney test: z=.29, p=.78). Hence, we pooled the results for the two texts. They also performed a comparable number of gestures in the first reading and in the second

reading (a mean of 32 in both cases: Wilcoxon test: z=07, p=.94). Hence, we pooled the results for the two readings.

As a general result, in the Gesture condition, the children followed the instruction to accompany the reading of the semantic units in the text with gestures. Twelve of the 24 children did so for all the 18 semantic units in the text, five children did so for 17 of the 18 semantic units, five for at least 14 of the semantic units, and two for at least 8 of the semantic units. Over the two readings in the Gesture condition, the children produced a mean of 50 representational gestures, six deictic gestures, six beat gestures, and one symbolic gesture.

Table 4 illustrates the mean percentages of types of recognition (Literal, Paraphrases, Wrong) as a function of the type of gesture accompanying at reading the semantic unit from which each triplet of sentences was created. In particular, for each participant we considered the percentage of times in which correct performance (accepting Literal, and rejecting Paraphrases and Wrong) and erroneous performance (rejecting Literal, and accepting Paraphrases and Wrong) was accompanied by each type of gesture while reading (at least one gesture of that type).

Insert Table 4 about here

An exploratory analysis based on such data revealed that literal sentences were more likely to be correctly accepted than refused when accompanied by representational gestures (Wilcoxon test: z=3.36, p=.001) and by beats (z=2.40, p<.02) while reading. Moreover, Paraphrases were more likely to be erroneously accepted than refuted when accompanied by representational gestures while reading (z=2.91, p=.004). As regards Wrong sentences, they were more likely to be correctly refuted than accepted when accompanied by representational gestures while reading (z=3.32, p=.001). However, as with the explorative analysis of gestures in Experiment 1, it was not possible to derive any strong conclusion from the emerging pattern given the lack of within-subject data.

Discussion

The results confirmed our prediction: when children gesticulated in the encoding phase they performed less well at rejecting Paraphrases than when they did not gesticulate: they tended to

confound the Paraphrases with the original sentences in the text. This result is consistent with the assumption that representing concepts through gestures in the encoding phase leads to the construction of an articulated mental model.

Interestingly, as far as the ability to recognize Literal sentences is concerned, we found no differences between performance in the two conditions. This result will require further investigation. Although we made no specific predictions about performance with Literal sentences, we hypothesized the presence of two competing cognitive processes: on the one hand, individuals who construct an articulated mental model tend to lose recall of verbatim text, on the other hand, individuals who enact concepts in the encoding phase tend to improve memory for verbatim text. Our experiment was not designed to disentangle these two possibilities; however our results seem to suggest that a certain type of interaction between contrasting effects is possible (and worth studying in future research).

General Discussion

In the present investigation we focused on the mechanisms underlying the facilitative effects of enactment on memory for texts. The results of Experiment 1 show that representing concepts through gestures in the learning phase leads to the construction of an articulated mental model of the text, as shown by the higher number of Correct recollections and Discourse-based inferences produced at recall. The results of Experiment 2 strengthen these findings; they show that representing concepts through gestures leads to the confounding of Paraphrases with the original sentences in the text, which is another index of the construction of an articulated mental model.

A weakness in our results is that the children in Experiment 1 drew few Discourse-based inferences, although all in the Gesture condition. It is possible that the technicality of the scientific texts prevented a stronger personal elaboration of the information in the text. This explanation would be consistent with the results of Experiment 2 which revealed that children were quite good

at recognizing text verbatim, suggesting a bias not to re-elaborate the text contents. Narrative texts might have favored more personal elaboration by the children participating in our experiments, thus resulting in the production of a greater number of Discourse-based inferences in Experiment 1 and a stronger bias to accept Paraphrases in Experiment 2. Further studies might attempt to extend our findings to learning from narrative texts, also in order to ascertain whether they can be generalized for contents other than scientific. Considered together, our results suggest that the mechanism underlying the beneficial effect of enactment on memory for text depends on a facilitating effect on the construction of an articulated mental model of the material to be learnt.

However, one may argue that poor verbatim memory in Experiment 2 can be accounted for by shifts of attention (see, e.g., Gernsbacher, 1985); in gesture condition, performing gestures might have distracted participants from memorizing surface forms. However, this explanation is not consistent with the results of Experiment 1, according to which recall was better in Gesture condition than in No-Gesture condition.

The literature on enactment offers several possible accounts for the facilitating effect of gesturing in the learning phase. A first attempted explanation, called the *multimodal account* (e.g., Engelkamp, 1998) emphasized the role of the motor components of subject-performed tasks (SPTs): performed actions enhance memory by incorporating a distinctive motor program into the memory trace. A similar account has been advanced by Mulligan and Hornstein (2003), who maintained that, in line with the transfer-appropriate-processing analysis of memory retrieval (e.g., Roediger & Guynn, 1996), "modality-specific information is encoded in memory for actions, in addition to verbal-semantic information" (ib., p. 419). An alternative explanation, known as the *episodic integration account* (e.g., Kormi-Nouri & Nilsson, 2001), claims that performed actions involve much richer and more elaborate representations than verbal phrases. As a consequence, the SPTs result in unitization of the object-action association: object and action components are encoded in a single memory unit, or in separate units with stronger interconnections (see also Mangels & Heinberg, 2006). Our proposed theoretical framework is consistent with such previous explanations,

although our proposal regards the beneficial effects of enactment on memory for text, whereas the abovementioned accounts regard the enactment effects on memory for sentences. Consider, for instance, the results in Engelkamp (1998), according to which in memory for sentences, enactment improves item-specific encoding but not relational processing. We interpret these results as confirmation of the difference between recall of semantically unrelated sentences and recall of sentences composing a text, that are semantically related. In the latter case, indeed, gestures may be just likely to improve item-specific encoding and favor richer and more elaborate representations, whereas in the former case gestures may exert their effect also favoring the construction of an articulated text mental model (that is not built in the case of enactment for sentences). According to our assumptions, the meaning of the sentences composing a text, as opposed to the meaning of semantically unrelated sentences, can be represented through an articulated mental model which supports Correct recollections and Discourse-based inferences at the expense of memory for text verbatim.

The literature on gestures also offers several accounts for the positive effect of gestures on learning. Three main lines of explanation for such an effect have been proposed: gesture may ground thought in action, they may bring new knowledge in to the individual's mental representation of the task, and gesturing may lighten cognitive load (see, e.g., Goldin-Meadow, 2010). According to the first line, gesturing introduces action information into mental representations and such action information then impacts on how individuals deal with the task they are performing. For instance, it has been shown that when gesturing adds action information to speakers' mental representations that are incompatible with subsequent actions, this information interferes with problem-solving, whereas when the information added by gestures is compatible with future actions, those actions will be facilitated (Goldin-Meadow & Beilock, 2013). It has also been shown that gesturing brings new knowledge into the learner's mental representation of the task: for instance, Broaders and colleagues (2007) found that children who were asked to gesture while solving mathematical equivalence problems expressed with gestures new and correct ideas

that they did not express in speech. According to the third line of explanation, gesturing while speaking has an impact on mental representations because it lightens the load on working memory. It has been shown that gesturing while performing explanation tasks, as compared to speaking without gesturing, reduces demands on the speaker's cognitive resources and frees up cognitive capacity to perform other tasks. For instance, Goldin-Meadow, Nusbaum and colleagues (2001) showed that gesturing on one task (explaining a math problem) affected performance on a second task (remembering a list of words or letters) in both adults and children.

The account we advances is in line with the former line of explanation: we assume that gesturing in the learning phase allows individuals to add information to their mental models . because gestures represent information in a spatial and non-discrete way, thus allowing such information to be easily inserted into the mental models under construction. More in general, our results extend to gestures the beneficial effects of types of visuo-spatial information other than diagrams on the construction and development of a text mental model (see, e.g. Bauer & Johnson-Laird, 1993; Butcher, 2006). Consistent, several studies suggest that gestures production also sustain spatial thinking, maybe because they sustain spatial representations in working memory (see, e.g., Bucciarelli, Khemlani *et al*., 2014; Morsella & Krauss, 2004). Second, it is possible that by facilitating the construction of a mental model, gestures also lighten the load on working memory, since the information is easily inserted into the mental model with less need for rehearsal

Still, one may argue that an alternative possible explanation to our results is that the Gesture task was simply more difficult than the No-Gesture task, and that such increased difficulty entail deeper processing and better text memory (see, e.g., the reverse coherence effect in O'Reilly & McNamara, 2007). If this would be the case, we should expect that in both recall and recognition tasks the performance of the participants in Gesture condition reflects a deeper processing, resulting in more Correct recollections and Discourse-based inferences in a recall task and a less memory for text verbatim in a recognition task. However, such tentative explanation contrasts with data in gesture literature showing that gesture production reduces cognitive burden, thereby freeing up

effort that can be allocated to other tasks, thus increasing the resources available (e.g., Goldin-Meadow, 1999). Therefore, at the state of the art, there is no consensus on the fact that gesture production, *per se,* makes a task more difficult.

In conclusion, consistent with the literature on text learning showing that active learning yields better results than passive reading, (e.g., Glenberg, Gutierrez *et al*., 2004; Noice & Noice, 2007),we argue that gesturing can be a proficient way to learn actively. Most important, our results confirm that gestures performed in the learning phase facilitate comprehension and learning even when they are produced on demand rather than spontaneously; the more important implication is that gestures can be fostered with the result of improving learning, also in children.

Acknowledgments

This work was supported by Italian Ministry of Education University and Research Grant 2010RP5RNM (to M.B.) to study problem solving and decision making.

References

Albrecht, J.E., & O'Brien, E.J. (1993). Updating a mental model: Maintaining both local and global coherence. *Journal of Experimental Psychology. Learning, Memory & Cognition; Vol. 19, Issue 5*, 1061-1070.

Alibali M.W., Evans J.L., Hostetter A.B., Ryan K., & Mainela-Arnold E. (2009) Gesture-speech integration in narrative discourse: Are children less redundant than adults? *Gesture, 9(1),* 290-311.

Bauer, M.I., & Johnson-Laird, P.N. (1993). How diagrams can improve reasoning. *Psychological Science, 4*, 372-378.

Bower, G. H., & Morrow, D. G. (1990). Mental models in narrative comprehension. *Science, 247(4938)*, 44-48.

Bransford, J. D., Barclay, J.R., & Franks, J. J. (1972). Sentence memory: A constructive versus interpretive approach. *Cognitive Psychology, 3*, 193-209.

Bransford, J. D., & Franks J.J. (1976). *Psychology of Learning and Motivation - Advances in Research and Theory, 10(C),* 93-127.

Broaders, S., Cook, S.W., Mitchell, Z., & Goldin-Meadow, S. (2007). Making children gesture brings out implicit knowledge and leads to learning. *Journal of Experimental Psychology: General, 136 (4)*, 539–550.

Bucciarelli, M., Khemlani, S., Mackiewicz, R., & Johnson-Laird, P.N. (2014). Gestures and simulations in children's algorithmic thinking. Under submission.

Butcher, K.R. (2006). Learning from text with diagrams: Promoting mental model development and inference generation. *Journal of Educational Psychology, 98*(1), 182-197.

Cohen, R.L. (1989). Memory for action events: The power of enactment. *Educational Psychology Review, 1*, 57-80.

Cohen, R. L., & Bryant, S. (1991). The role of duration in memory and metamemory of enacted instructions (SPTs). *Psychological research, 53(3)*, 183–187.

Colletta, J. M. (2009). Comparative analysis of children's narratives at different ages: A multimodal approach. *Gesture*, *9*(1), 61-96.

Colletta, J.M., Pellenq, C. & Guidetti, M. (2010). Age related changes in co-speech gestures and narratives : evidence from French children and adults. *Speech Communication*, *52*, 565-576.

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

Cutica, I. & Bucciarelli, M. (2011). "The more your gestured, the less I gesture": Co-speech gestures as a measure of mental model quality. *Journal of Nonverbal Behavior*, *35*, 173-187.

Cutica, I. e 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.

Engelkamp, J. (1998). *Memory for actions*. Hove: U.K.: Psychology Press.

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

Feyereisen, P. (2009). Enactment effects and integration processes in younger and older adults' memory for actions. *Memory*, *17*, 374-385.

Frick-Horbury, D. (2002). The use of hand gestures as self-generated cues for recall of verbally associated targets. *American Journal of Psychology, 115*, 1-20.

Gernsbacher, M.A. (1985). Surface information loss in comprehension. *Cognitive Psychology, 17*, 324-363.

Glenberg, A.M., Gutierrez, T., Levin, J.R., Japuntick, S., & Kaschak, M.P. (2004). Activity and imagined activity can enhance young children's reading comprehension. *Journal of Educational Psychology, 96*(3), 424-436.

Glenberg, A. M., Kruley, P., & Langston, W. E. (1994). Analogical processes in comprehension Simulation of a mental model. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics*(pp. 609–640). New York: Academic Press.

Glenberg, A. M., Meyer, M., & Lindem, K. (1987). Mental models contribute to foregrounding during text comprehension. *Journal of Memory and Language*, *26*(1), 69-83.

Goldin-Meadow, S. (1999). The role of gesture in communication and thinking. *Trends in cognitive sciences*, *3*(11), 419-429.

Goldin-Meadow, S. (2010). When gesture does and does not promote learning. *Language and Cognition*, *2*(1), 1-19.

Goldin-Meadow, S. & Beilock, S.L. (2013). Action's influence on thought: The case of gesture. *Psychological Science, 5*(6), 664-674.

Goldin-Meadow, S., Cook, S.W., & Mitchell, Z.A. (2009). Gesturing gives children new ideas about math. *Psychological Science, 20 (3)*, 267–272.

Goldin-Meadow, S., Kim, S., & Singer, M. (1999). What the teacher's hands tell the student's mind about math. *Journal of Educational Psychology, 91*, 720-730.

Goldin-Meadow, S., Levine, S. C., Zinchenko, E., Yip, T. K., Hemani, N., & Factor, L. (2012). Doing gesture promotes learning a mental transformation task better than seeing gesture. *Developmental science*, *15*(6), 876-884.

Goldin-Meadow, S., Nusbaum, H., Kelly, S., & Wagner, S. (2001). Explaining math: Gesturing lightens the load. *Psychological Science, 12*, 516-522.

Graesser, A. C., Gernsbacher, M. A., & Goldman, S. R. (Eds.). (2012). *Handbook of discourse processes*. Routledge.

Grasser A. C., Millis K. K., & Zwaan R.A. (1997). Discourse comprehension. *Annual Review of Psychology, 48*, 163-189.

Graesser, A. C., Singer M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review, 101*, 371–395.

Jahn, G. (2004). Three turtles in danger: Spontaneous construction of causally relevant spatial situation models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(5), 969.

Johnson-Laird, P.N. (1983). *Mental models: Towards a cognitive science of language, and consciousness*. Cambridge University Press, Cambridge, UK.

Johnson-Laird, P.N. (2006). *How we reason*. New York: Oxford University Press.

Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Johnson-Laird, P. N., & Stevenson, R. (1970). Memory for syntax. *Nature*, *227*, 412.

Kaup, B., Kelter, S., & Habel, C. (1999). Taking the functional aspect of mental models as a starting point for studying discourse comprehension. In G. Rickheit & C. Habel (Eds.), *Mental models in discourse processing and reasoning* (pp. 93 –112). Amsterdam: North-Holland.

Kelly, S.D., Manning, S., & Rodak, S. (2008). Gesture gives a hand to language and learning: Perspectives from cognitive neuroscience, developmental psychology and education. *Language and Linguistics Compass*, *2*, 1–20.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge University Press, USA.

Knauff, M. & Johnson-Laird, P.N. (2002). Visual imagery can impede reasoning. *Memory & Cognition, 30*(3), 363-371.

Kormi-Nouri, R., & Nilsson, L-G. (2001). The motor component is not crucial! In H.D. Zimmer, R.L. Cohen, M.J. Guynn, J. Engelkamp, R. Kormi-Nouri & M.A. Foley (Eds.), *Memory for action: A distinct form of episodic memory?* (pp. 97-111). New York: Oxford University Press.

Kunene, R., & Colletta, J. M. (2007). Exploration of multimodal narratives: a cross-language comparison of isiZulu and French. In *Proceedings of the 10th International Pragmatics Conference, Göteborg, July* (pp. 8-13).

Mangels, J.A., & Heinberg, A. (2006). Improved episodic integration through enactment: Implications for aging. *The Journal of General Psychology, 133*, 37-65.

Mani, K., & Johnson-Laird, P. N. (1982). The mental representation of spatial descriptions. *Memory & Cognition, 10*, 181-187.

McNamara, T., Miller, D.L., & Bransford, J.D. (1991). Mental models and reading comprehension. R. Barr, M.L. Kamil, P.B. Mosenthal, P.D. Pearson (Eds.), *Handbook of Reading Research, Vol. 2*, 490-511. Erlbaum, Hillsdale.

McNeill, D. (1992). *Hand and Mind*. Chicago: University of Chicago Press.

Morsella, E., & Krauss, R.M. (2004). The role of gestures in spatial working memory and speech. *American Journal of Psychology, 117*(3), 411-424.

Mulligan, N.W., & Hornstein, S.L. (2003). Memory for actions: Self-performed tasks and the reenactment effect. *Memory & Cognition, 31*, 412-421.

Noice, H., & Noice, T. (2001). Learning dialogue with and without movement. *Memory & Cognition, 29*, 820-827.

Noice, H., & Noice, T. (2007). The non-literal enactment effect: Filling in the blanks. *Discourse Processes, 44*, 73-89.

O'Reilly, T., & McNamara, D. S. (2007). Reversing the reverse cohesion effect: Good texts can be better for strategic, high-knowledge readers. *Discourse Processes, 43*, 121–152.

Ping, R.M., & Goldin-Meadow, S. (2008). Hands in the air: Using ungrounded iconic gestures to teach children conservation of quantity. *Developmental Psychology, 44(5)*, 1277-1287.

Roediger, H. L., & Guynn, M. J. (1996). Retrieval processes. In E. L. Bjork, & R. A. Bjork (Eds.), *Memory* (pp. 197–236). New York: Academic Press.

Singer, M. (1994). Discourse inference processes. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 479–515). San Diego, CA: Academic.

van Dijk, I.A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.

Vendrame, M., Cutica, I., & Bucciarelli, M. (2010). “I see what you mean”: Oral deaf individuals benefit from speaker’s gesturing. *The European Journal of Cognitive Psychology, 22*, 4, 612-639.

von Essen, J.D. (2005). Enactment enhances integration between verb and noun, but not relational processing, in episodic memory. *Scandinavian Journal of Psychology, 46*, 315-321.

Wagner Cook, & Goldin-Meadow, S. (2006). The role of gestures in learning: Do children use their hands to change their minds? *Journal of Cognition and Development, 7*(2), 211-232.

Wagner Cook, S., Mitchell, Z., & Goldin-Meadow, S. (2008). Gesturing makes learning last. *Cognition, 106*, 1047-1058.

Zimmer, H.D. (2001). Why do actions speak louder than words? Action memory as a variant of encoding manipulation or the result of a specific memory system? In H.D. Zimmer, R.L. Cohen, M.J. Guynn & J. Engelkamp (Eds.), *Memory for action: A distinct form of episodic memory?* (pp. 151-198). New York: Oxford University Press.

Zwaan, R. A., Magliano J. P., & Graesser, A C. (1995). Dimensions of situation model construction in narrative comprehension. *Journal of Experimental Psychology: Learning, Memory and Cognition, 21*, 386-397.

Zwaan R. A., & Radvansky G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin, 123(2),* 162-185.

Table 1. Mean types of recollections in the two conditions of Experiment 1.

Condition	Literal recollections	Correct recollections	Discourse-based inferences	Elaborative inferences	Errors
Gesture (N=24)					
M	0.4	6.6	0.3	0.1	0.4
SD	0.7	3.0	0.7	0.3	0.7
No-Gesture (N=24)					
M	0.4	5.2	0.0	0.1	1.3
SD	0.6	2.8	0.2	0.3	1.6

Table 2. Mean percentages of types of recollection (Literal, Proper and Wrong) as a function of the type of gesture at reading in Experiment 1.

Type of gesture	**Type of recollection**		
at reading	Literal	Proper (Correct, Proper)	*Wrong* *(Elaborative, Errors)*
Representational	*14*	*51*	*20*
Deictic	*13*	*18*	*8*
Beats	*0*	*15*	*13*
Symbolic	*0*	*2*	*0*

Table 3. Mean of types of sentences accepted in the two experimental conditions in Experiment 2.

Condition	Paraphrases (n=12)	Literally correct (n=12)	Wrong content (n=12)
Gesture (N=24)			
M	2.5	4.7	0.9
SD	1.5	1.3	0.9
No-Gesture (N=24)			
M	1.8	4.6	1.0
SD	1.4	1.0	0.9

Table 4. Mean percentages of Literal, Paraphrases and Wrong content sentences accepted and refused, as a function of the type of gesture at reading in the Gesture condition. For example, the table shows that the 56% of Literal sentences accepted were accompanied by representational gestures at reading.

Type of gesture at reading	**Type of recollection**					
	Literal		Paraphrases		*Wrong*	
	ACC(*)	RIF	ACC	RIF(*)	ACC	RIF(*)
Representational	56	31	53	31	25	53
Deictic	22	19	26	22	13	22
Beats	19	2	9	22	17	17
Symbolic	1	1	1	1	0	2

Appendix A

A.1 The circulatory system text used in both Experiments 1 and 2

(Semantic units are separated by slashes)

The circulatory system allows blood to flow,/ in order to transport nutrients and oxygen throughout the body/ and collect the waste produced by the cells./ Blood flows through a complex system of blood vessels,/ made up of arteries, veins and capillaries./ Blood is made up of a liquid part, called plasma,/ and various types of cells that have different functions: red blood cells, white blood cells, platelets./ Plasma is a yellowish liquid; it consists mainly of water, in which both the nutrients and waste are dissolved./ The heart is the motor of the circulatory system, / it works constantly to keep the blood moving through all the blood vessels./ The heart is a hollow, involuntary muscle:/ it is about the size of a fist and is located in the chest, between the two lungs. / Movements of the heart are called pulsations or heart beats./ Internally the heart is divided into two sides, which are not communicating./ Only oxygen-rich blood flows through the left side of the heart/ while blood rich in carbon dioxide flows through the right side./ Each side is divided into an upper cavity, called the atrium,/ and a lower cavity, called the ventricle.

A. 2 Examples of sentences used for the Recognition Task of Experiment 2

Literal	Blood transports nutrients and oxygen throughout the body
Paraphrases	Blood carries oxygen and nutrients to the cells in the body
Wrong	Blood transports waste products to the cells in the body

Literal	Movements of the heart are called pulsations or heart beats
Paraphrases	Contractions of the heart are called heart beats or pulsations
Wrong	Movements of the heart are called venous pulsations or impulses

Appendix B

B.1 The pulling force text used in both Experiments 1 and 2

(Semantic units are separated by slashes)

In nature, all objects exert a force/ of attraction on anything that is nearby./ This force is called force of gravity./ For example, a pen attracts the eraser towards it and vice versa,/ but the force of attraction is so small that it has no effect./ On the other hand, the force exerted by the Earth is extremely strong,/ so much so that everything is attracted towards the centre of the Earth:/ the atmosphere, the oceans, houses, cars, trees, animals and even human beings./ It is due to the force of gravity that when we lose our balance, we fall to the ground./ The force of gravity exerted by the Earth is so strong that even the Moon is attracted towards it,/ and that is why it revolves around the Earth without ever being able to pull away./ The pulling force exerted by the Moon on the Earth is, instead, much smaller/ because the mass of the Moon is much smaller than that of our planet./ The Earth is, in turn, attracted to the Sun,/ around which it moves in a circular orbit./ The same happens for all the other planets and stars in the universe:/ each one is attracted towards its closest neighbor/ and so revolves around it.

B. 2 Examples of sentences used for the Recognition Task of Experiment 2

Literal	For example, a pen attracts the eraser towards it and vice versa
Paraphrases	To give an example, an eraser and a pen attract one another
Wrong	For example, a pen attracts an eraser towards it, but not vice versa
Literal	The pulling force exerted by the Moon on the Earth is, instead, much smaller
Paraphrases	The force of gravity exerted by the Moon is smaller than that of the Earth
Wrong	The pulling force exerted by the Moon on the Earth is, instead, much greater