"The more you gestured, the less I gesture": Co-speech gestures as a measure of mental model quality.

"The more you Gestured, the less I Gesture":

Co-speech Gestures as a Measure of Mental Model Quality

Ilaria Cutica*, Monica Bucciarelli*

*Center for Cognitive Science, Department of Psychology

University of Turin

Turin, Italy

Corresponding author: Ilaria Cutica, Ph.D

Centro di Scienza Cognitiva and Dipartimento di Psicologia,

University of Turin, Via Po 14, 10123 Turin, Italy.

Phone: +39 011 670 30 74 Fax +39 011 8159039

Email: ilaria.cutica@unito.it

Abstract

Studies in the literature have revealed that a speaker's co-speech gestures favor the construction of a complete and articulated mental model of the discourse by the listener; moreover, from the literature on co-speech gestures we know that they help the speaker to organize the stream of thought. Given these data, we hypothesized that a person who listens to a discourse accompanied by gestures, as compared to a person who listens to a discourse not accompanied by gestures, would produce fewer co-speech gestures in recollecting the discourse. The analysis of the co-speech gestures produced by the participants in two experiments while recollecting the content of a discourse confirmed our predictions.

Keywords**:** co-speech gestures; discourse production; discourse comprehension; mental models

Introduction

Numerous studies in the literature have shown that deep comprehension of a discourse involves the construction and manipulation of mental representations that reproduce the state of affairs described (Graesser, Millis, Zwaan, 1997; McNamara, Miller & Bransford, 1991; Zwaan, Magliano & Graesser, 1995; Zwaan & Radvansky, 1998). The listener builds such representations on the basis of the semantic and pragmatic information contained in the text, together with his or her prior knowledge and any inferences that are drawn. Depending on the different theoretical frameworks within which this phenomenon has been studied, these mental representations are called mental models (Johnson-Laird, 1983; 2006) or situational models (van Dijk & Kintsch, 1983; Kintsch, 1998). A mental model is a mental representation that analogically reproduces a perceived or described state of affairs; it consists of elements, which stand for the perceived or described entities, and relationships between such elements, which stand for relationships between the entities. Several authors, including Biggs (2001), have demonstrated that deep comprehension and learning require high-level cognitive activity, unlike superficial comprehension and mnemonic learning. Within the theoretical paradigm of mental models, such high-level cognitive activities consist in the construction and manipulation of mental models. The more a listener is able to make links and place the information that is received within an integrated network, the higher his level of comprehension.

It has also been demonstrated that co-speech gestures have a positive effect on the listener's comprehension of a discourse. These gestures have been found to provide information that interacts with that conveyed by speech to reconstruct the communicative meaning (see e.g., Alibali, Flevares & Golden-Meadow, 1997; Goldin-Meadow, 1999; Kelly & Church, 1998). One possible explanation for this phenomenon is that the information conveyed by the speaker's co-speech gestures, represented in a non-discrete format, is easily

included in the mental model of the discourse (Bucciarelli, 2007, Cutica & Bucciarelli, 2008), since mental models too are non-discrete representations (see Hildebrandt, Moratz, Rickheit & Sagerer, 1999; Rickheit & Sicheleschmidt, 1999). In other words, co-speech gestures might lead to the construction of representations that are easily incorporated into the discourse model, alongside the representations constructed on the basis of the verbal information, enriching these and completing the mental model.

On the other hand, numerous findings in the literature have suggested that gesticulation is also involved in the speaker's mental organization of the discourse by helping to organize the stream of thought (see e.g., Alibali, Kita & Young, 2000; Kendon, 1983; McNeill, 1992; Goldin-Meadow & Alibali, 1999). Gestures accompany speech in the majority of conversations (McNeill, 1992), across almost all human cultures (Cassell, 1998); words and gestures can be seen as aspects of the same communicative event (Goodwin, 1986), generated by a single semantic representation that the speaker intends to communicate (McNeill, 1992). It has been widely observed that people gesticulate even when the listener cannot see them, for instance during telephone conversations, or when a screen deprives them of reciprocal visibility (see, for example, Rimé, 1982). Even people who are blind from birth and have thus no experience with the communicative function of gesture, produce gestures that resemble sighted-people's gestures in both form and content (Iverson & Goldin-Meadow, 1998). McNeill (1992) hypothesized that, in the speaker, words and gestures originate from a common mental representation, a kind of seed out of which they both develop, and that the message they both represent is divided into the two channels, speech and gesture, as the thought is expressed. Consistently, Cook, Jaeger and Tanenhaus (1999) found that speakers gesture more and are less fluent as discourse production increases in difficulty. It has also been demonstrated that gesturing can facilitate thought, reducing the cognitive load and leaving more resources available for other tasks (Alibali & DiRusso, 1999). Various

hypotheses have been proposed to explain how gestures can facilitate verbal expression of concepts; one possible explanation is that gestures function at the level of generating the surface form of the utterance, specifically by facilitating access to words in the mental lexicon (see, for example, Butterworth & Hadar, 1989; Krauss, Chen & Chawla, 1996; Rauscher, Krauss & Chen, 1996). According to an alternative explanation, backed by more experimental evidence than other current theories, gestures might facilitate the organization of information at a "higher" level than speech production: they might be helpful in the conceptual planning of utterances (see, for example, Alibali et al., 2000, Kita, 2000). It has, for instance, been demonstrated that the action of gesturing helps people focus on perceptual and motor knowledge: when prevented from gesturing, people provide less perceptual and motor information when describing an event than when they are able to gesture (Alibali et al., 2000). Likewise, the imagery content of a guided conversation is reduced when gesturing is prevented (Gyselinck & Tardieu, 1999).

Considering the above results as a whole, we hypothesized that speakers who have a good mental model (i.e., a complete and articulated mental model) of their discourse have less need to produce gestures when proffering a discourse (to organize their thoughts) than speakers who have a poorer mental model of the discourse.

How Gestures Contribute to the Construction of the Mental Model of a Discourse

The spontaneous gestures that accompany speech are also referred to as non-symbolic gestures. These have been classified in different ways in the literature; however, most authors generally agree in distinguishing three main categories:

(1) *deictic gestures*, consist of indicative or pointing acts, and are commonly used to address 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);

(2) *representational gestures* (McNeill, Cassell & McCollough, 1994), or *ideational gestures* (Hadar, Burnstein, Krauss & Soroker, 1998), or *illustrator gestures* (Ekman & Friesen, 1972), which refer to actions, characteristics, relationships between objects or people, forms or movements, whether concrete (*iconic gestures)*, or metaphoric (*metaphoric gestures*), which can be redundant or add meaning to the discourse;

(3) *batons* (Efron, 1941), or *beats* (Kendon, 1983; McNeill, 1987), or *motor gestures* (Krauss, Chen & Gottesman, 2000), rhythmic or repetitive movements that are not related to the semantic content of the accompanying words or sentences (Feyereisen, Van de Wiele & Dubois, 1988), but are coordinated with the speech prosody and fall on stressed syllables (Bull & Connelly, 1985).

For our purposes, we distinguished between deictic and representational gestures, on the one hand, and beat gestures, on the other, based on the presence or otherwise of a relationship between a gesture and the semantic content of the co-occurring speech. Although we had no specific predictions, we were interested in investigating whether deictic and representational gestures, which we shall refer to as “semantic gestures”, have a different effect on comprehension and memory than beat gestures.

Although co-speech gestures, when considered as a whole, have positive effects on comprehension, it has been demonstrated that different types of gestures can produce different communicative effects on the listener. For example, only deictic gestures are helpful for understanding the real communicative intentions that underlie an indirect request (Kelly, Barr, Church & Lynch, 1999). As regards representational gestures, some studies have demonstrated that both children (Kelly & Church, 1997) and adults (Goldin-Meadow, Wein & Chang, 1992) are able to decode and use information conveyed through these, and that they facilitate recollection of utterances more than others (Thompson, Driscoll & Markson, 1998).

Moreover, when recollecting a discourse people include the meaning of representational gestures even when the task only requires them to recall the contents of the verbal part of the discourse (Thompson et al., 1998).

Recent studies have suggested that gestures accompanying speech might, more generally, help the listener to construct a complete and articulated mental model of its contents (Bucciarelli, 2007; Cutica & Bucciarelli, 2008). Indeed, a person listening to a discourse by a gesticulating speaker remembers more information and draws more discourse-based inferences than a person listening to the same discourse by a non-gesticulating speaker. These two indexes (the amount of correct information recollected and the number of inferences drawn from information explicitly stated in the discourse) are relevant because they are reliable indicators of the construction of a good mental model of a discourse (Bransford, Barclay & Franks, 1972; Johnson-Laird, 1983; Johnson-Laird & Byrne, 1991). For our purposes, we distinguished between discourse-based inferences and elaborative inferences, arguing that 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 (i.e., beliefs and intentions) with respect to the actions described (Graesser, Singer & Trabasso, 1994). Elaborative inferences (e.g., Singer, 1994) are instead a sort of enrichment of the original text. Given these definitions, discourse-based inferences are the only type of inferences that indicate the construction of a mental model of the discourse.

Gestures as a Measure of Mental Model Quality

Data in the literature support the assumption that speakers who have a good mental model of the contents of a discourse have less need to produce gestures when proffering the discourse than speakers who do not have such a good mental model of their discourse. On the other hand, we know that a person listening to a discourse accompanied by gestures

constructs a better mental model than a person listening to the same discourse not accompanied by gestures. Therefore, we would expect an individual who has listened to a discourse accompanied by gestures, and has thus constructed a good mental model, to produce fewer gestures, when recalling the contents of the discourse, than an individual who has listened to the same discourse by a non-gesticulating speaker.

Moreover, since the literature has shown that different types of gestures can specifically influence the listener's comprehension of the discourse, we asked whether and how this difference might affect the quality of the listener's mental model. We made a distinction between gestures relating to the semantic content of the discourse (representational gestures and deictic gestures) and those relating to the pragmatic content (beats), and investigated whether one type affected the construction of the mental model of the discourse more than the other. For that purpose, we conducted two types of analyses of the speech and gestures produced by listeners when recollecting the discourse. First, based on the assumption that, as reported in the literature, semantic gestures facilitate discourse comprehension more than beats, we expected listeners to properly recollect concepts (i.e., in the form of correct recollections and discourse-based inferences) which the speaker had accompanied with semantic gestures better than concepts accompanied by beats (as they should have a better mental model of the concepts accompanied by the speaker's semantic gestures). Second, we expected listeners to produce fewer gestures when properly recollecting concepts which the speaker had accompanied with semantic gestures than when recollecting those accompanied by beats (having a better mental model of the units accompanied by the speaker's semantic gestures requires less cognitive effort to recollect them).

To verify our predictions we analyzed the recollections and co-speech gestures produced by the participants in two experiments conducted by Cutica and Bucciarelli (2008).

Participants

Experiment 1 Thirty-eight students (32 females and 6 males, mean age: 23 years) at Turin University took part as a group in the experiment, which earned them course credits.

Experiment 2 Thirty students (24 females and 6 males, mean age: 23 years) at Turin University took part as a group in the experiment, which earned them course credits. None of the participants in Experiment 2 took part in Experiment 1.

Material and Procedures

Experiment 1 The experimental material for the first experiment consisted of two films, each lasting approximately six minutes, in which an actor described a series of events that took place at a funfair/carnival (see Appendix 1 for an excerpt): this was a narrative text. The actor proffered the same discourse in both films, but in two different conditions; in one film the discourse was accompanied by the actor's spontaneous gestures (he was instructed to produce hand and arm movements as he felt appropriate with respect to the discourse flow, *Gesture condition*), in the other it was not accompanied by any gestures (*No Gesture condition*). In the Gesture condition, the actor accompanied each of the semantic units that we identified with one or more gestures. The actor was filmed sitting down in both conditions, so that he did not appear unnaturally immobile. Half of the participants (balanced for age and gender) were assigned to the Gesture condition, and half to the No Gesture condition. The participants were invited to watch one of the two films carefully and to pay maximum attention to the words spoken by the actor. At the end of the film, the participants were asked to move to another position, where they could be video-recorded, and were asked to recall what the actor had said, in as much detail as possible.

In order to codify the participants' recollections, the discourse was divided into 54 semantic units which corresponded to the single concepts it was possible for them to recall (the division into semantic units is shown in Appendix 1). Each concept (i.e., semantic unit) recalled by the participants was evaluated according to the following coding schema:

- Correct recollection: a semantic unit recollected either in its literality or 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: a semantic unit recollected with the addition of plausible details.
- Erroneous recollection: a recollection with a meaning that was inconsistent with the semantic unit.

Consider, for instance, the following semantic unit in the discourse: “Night was falling, and the shadows were growing longer”. According to the coding schema, the statement “It was beginning to grow dark” is a correct recollection; the statement “Since it was getting dark, the search became more difficult” is a discourse-based inference (because it refers to a causal consequent); the statement “Suddenly everything [*the roundabouts*] closed down, and the shadows were growing longer” is an erroneous recollection. Now consider the following semantic unit in the discourse: “She was clinging all alone to an apocalyptical beast”; according to the coding schema, “She was on the roundabout, her arms around a golden-headed monster” is an elaborative inference.

We are entitled to exclude differences, other than co-speech gestures, between the two conditions.. Indeed, the speaker was a professional actor who learnt the discourse by heart, and he was instructed to produce the same facial expressions and the same voice intonation in both conditions. However, as concerning the prosody and the intonation of voice, Cutica and Bucciarelli (2008) carried out a pre-test on the audio recording of the two conditions. Half of the participants in the pre-test were presented with the audio recording pertaining to the fiction in the Gesture condition, and half with the audio recording of the fiction pertaining to the No Gesture condition. They were invited to listen to the audio

recording, and as soon as it ended they were asked to recall as much information as they could. All of them were video-recorded. The results showed no difference between performance by the participants assigned to the audio recording of the Gesture condition and those assigned to the audio recording of the No Gesture condition: indeed, they did not differ as regards the number of correct recollections, discourse-based inferences, elaborative inferences, and erroneous recollections. Thus, there was no difference in the two audio recordings pertaining to the different experimental conditions.

Experiment 2 Experiment 1 was based on the use of a narrative discourse with high spatial and movement content. Since some studies have revealed a particular link between spontaneous gesture production and contents involving visual or motor imagery, i.e., spatial and movement content (e.g., Alibali et al., 2001; Butterworth & Hadar, 1989; Feyereisen & Harvard, 1999; Krauss, 1998; Krauss et al., 2000; Rauscher et al., 1996), the effects observed in Experiment 1 may have been influenced by the specific type of content that was chosen: co-speech gestures may be more effective with high spatial content than with low spatial content. Thus, Cutica and Bucciarelli (2008) replicated Experiment 1, using a discourse with low spatial and movement content. Within the advanced theoretical framework, since the construction of a good mental model of the discourse should be independent of its degree of spatial content, the speaker's spontaneous gestures should also facilitate the listener’s comprehension and recollection of discourses with low spatial content. (However, discourses with a high spatial and motion content might still benefit from gestures more than those with low spatial and motion content; a possibility not explored by the present investigation). The second experiment was carried out in exactly the same way as the first, but the actor proffered a technical discourse with low spatial and movement content, on color perception (see Appendix 2 for an excerpt). As for Experiment 1, the discourse was accompanied by gestures in the Gesture condition, but not in the No Gesture condition. Half of the participants

(balanced by age and gender) were assigned to the Gesture condition and half to the No Gesture condition. As in the previous experiment, in order to codify their recollections, the discourse was divided into semantic units; 35 semantic units were identified, which corresponded to the single concepts it was possible for the participants to recall (the division into semantic units is shown in Appendix 2). The coding schema applied to this discourse was the same as the one used for Experiment 1. For instance, the discourse contained the following sentence: “It’s beyond dispute that colors carry strong expressive components”; the statement “It is well known that color is a powerful expressive means” is a correct recollection; the statement “Several researchers have demonstrated that color is an important expressive means” is a discourse-based inference (because it refers to a causal antecedent), and the statement “The expressive characteristics and functions of color are the subject of much debate” is an erroneous recollection. Now consider the following statement in the original discourse: “Red should be considered exciting because it reminds us of the connotations of fire, blood and revolution”; the statement “Red can evoke passion, blood and revolution” is an elaborative inference.

The actor’s facial expressions were the same in the films of the Gesture and No Gesture conditions. Further, as for the material used in Experiments 1, Cutica and Bucciarelli (2008) carried out a pre-test to compare the audio recording of the Gesture condition with the audio recording of the No Gesture condition. The results revealed no difference in performance between the participants assigned to the two conditions, as evaluated through correct recollections, discourse-based inferences, elaborative inferences and erroneous recollections.

Results

As stated above, to verify the predictions about the relationship between the gestures produced by an actor and the gestures produced by the listener in the recollection phase, we

analyzed the quantity and the type of co-speech gestures produced by the participants in the experiments by Cutica and Bucciarelli (2008), in relationship to their verbal recollections. For the sake of clarity, we shall now briefly report the results of our previous study, in which we analized the participants' verbal recollections. In that work, we compared participants' recollections (whether or not they were accompanied by gestures) in the Gesture and in the No Gesture conditions, and found that a discourse accompanied by gestures, as compared with a discourse not accompanied by gestures, resulted in better recollection of conceptual information, a greater number of discourse-based inferences, and poorer recognition of verbatim of the discourse[1]. However, we found no differences in the number of erroneous recollections produced by the participants: mental models do not prevent a person from making mistakes. It may depend on the fact that if the listener misunderstands some piece of information, there is a chance that the misunderstood information will be included in the mental model, thus supporting a wrong recollection. A good mental model may avoid this possibility only when the misunderstood information is in conflict with some other piece of information (a good mental model is fully coherent).

The focus of the present investigation was, instead, the analysis of co-speech gestures produced by participants in the recollection phase. We defined as gestures accompanying a semantic unit those hand movements that had a clear beginning and an end point, and that were also temporally linked to the proffering of the utterance relating to the semantic unit. An initial analysis of the participants' gestures when recollecting each semantic unit excluded the possibility that they were repeating the same gestures produced by the actor

[1] Another possible indicator of the building of an articulated mental model of a given text material is the fact that the listener has a poor retention of the surface form of the text (for experimental evidence see Johnson-Laird & Stevenson, 1970; Garnham, Oakhill, Cain, 1998). This happens because, according to the mental model theory, mental models derived from a text do not encode the linguistic form of the sentences on which they are based. The results reported by Cutica and Bucciarelli (2008) confirmed that people listening to a discourse accompanied by gestures recall less information at verbatim level than those listening to a discourse not accompanied by gestures. We have not addressed this issue in the present paper, as verbatim is better evaluated through recognition tasks, and recognition tasks are not suitable to investigate co-speech gestures.

while proffering that specific semantic unit.

Results for Experiment 1

Three independent judges coded the participants' verbal recollections of the 54 semantic units individually. They reached a significant level of agreement on their first judgments (Cohen's K ranging from .92 to .94, all p<.001). For the final score, the judges discussed each item on which they disagreed, until reaching a full agreement. Two more independent judges examined the gestures produced by each participant and initially agreed to recognize 89% of their hand movements as gestures; the level of agreement was significant (Cohen's K ranging from .93 to .94, all p <.001). For the final score, the judges discussed each item on which they disagreed, until reaching a full agreement. By way of example, the following movement was considered to be an accompanying gesture: raising the right hand until it was level with the mouth, with the palm towards the face and the fingers spread out, and then moving it forward quickly while saying the words "and he blew on the candy floss" (correct recollection). Table 1 shows the results according to the type of recollection.

The results confirmed our predictions: in the recollection phase, participants in the Gesture condition produced fewer co-speech gestures (41% of recollections accompanied by gestures: 10.05 units accompanied by gestures out of 24.43 recollected units) than participants in the No Gesture condition (56% of recollections accompanied by gestures: 9.32 units accompanied by gestures out of 16.79 recollected units; t-test for independent samples: t(36)=-2.70, tied p=011). More in detail, as regards correct recollections, participants in the Gesture condition accompanied fewer recollected units with gestures than those in the No Gesture condition (t-test for independent samples: t(36)=-2.71, tied p=.01). As regards errors, discourse-based inferences, and elaborative inferences, there was no significant difference in the production of co-speech gestures by the participants in the two conditions (t-test for independent samples: t(36) ranging from -.92 to 0, p ranging from .37 to 1).

Next we investigated whether the semantic gestures and beats produced by the actor in the Gesture condition had different effects on the building of a good mental model of the discourse, namely in the production of correct recollections and of discourse-based inferences by the listener. For that purpose, we only considered the semantic units that the actor had accompanied with a single type of gesture, excluding those semantic units that he had accompanied with a mixture of gestures (for example, a beat-iconic gesture sequence). We thus obtained 33 semantic units accompanied by "pure" gestures: 21 semantic gestures (16 representational gestures, 5 deictic gestures) and 12 beats. Table 2 shows the mean number of proper recollections (that is, correct recollections + discourse-based inferences) according to the dichotomy beats-semantic gestures.

Given the different number of gestures produced by the actor in each category, we performed our statistical analysis by considering, for each participant, the number of proper recollections that referred to discourse units accompanied by a certain type of gesture performed by the actor, with respect to the total number of discourse units that the actor had accompanied with that type of gesture. The results revealed that the actor's semantic gestures facilitated proper recollections significantly better than beats (paired t-test, $t(18)=2.40$, $p=.02$). A more detailed study within the semantic gestures category, involving an analysis of variance, revealed that proper recollections varied in relation to the gestures with which the actor had accompanied the respective semantic units (ANOVA: $F_{3,29}=4.30$, $p=.01$); in particular, metaphoric gestures and deictic gestures were associated with proper recollections more than iconic gestures (Scheffe: $p =.02$).

We also predicted that listeners would recollect concepts which the speaker had accompanied with semantic gestures better than those accompanied by beats, as they should have a better mental model of the concepts accompanied by the speaker's semantic gestures. We conducted our statistical analysis by considering, for each participant, the number of

gestures produced when properly recalling (in the form of correct recollection or discourse-based inference) a discourse unit that the actor had accompanied with a certain type of gesture, with respect to the total number of discourse units that the actor accompanied with that type of gesture. Listeners produced a mean of 1.5 (sd=1.75) accompanying gestures when properly recollecting the 21 semantic units that the actor had accompanied with one or more semantic gestures (either representational or deictic), while they produced a mean of 3.27 (sd=2.31) accompanying gestures when recollecting the 12 semantic units that the actor had accompanied with one or more beats. The results revealed that, in the recollection stage, listeners produced more gestures when reporting concepts that the actor had accompanied with beats than when reporting concepts accompanied by semantic gestures (paired t-test, $t(18)=2.31$ $p=.03$). Given the small number of gestures produced by the participants, it was not possible to carry out a statistical analysis of the representational gestures and deictic gestures separately.

If the actor used semantic gestures with content that was intrinsically easier to remember (perhaps very easy to visualize such as words with spatial content) then the listeners' recollections might have been better and this in turn might have influenced their use of gestures. In this case, the influential factor is the verbal content uttered by the actor, not the gestures he used. In order to exclude such a possible explanation for our results, we examined the proper recollections by participants in the No Gesture condition of the same 33 units that, in the gesture condition, the actor had accompanied with a "pure" gesture. Obviously, in the No Gesture condition, those units were not accompanied by any gesture, so that the only difference between them was the semantic content of the speech. We thus compared, in the No gesture condition, the number of proper recollections of the 21 units that, in the Gesture condition, were accompanied by semantic gestures, with the number of proper recollections of the 12 units that, in the Gesture condition, were accompanied by beats. We found no

difference between participants' proper recollections (paired t-test, $t(18)=.532$, $p=.61$). This finding excludes the possibility that the influential factor is the verbal content uttered by the actor, rather than the gestures he used.

Results for Experiment 2

Three independent judges coded the participants' verbal recollections individually, considering the 35 semantic units into which the discourse was divided; they reached a significant level of agreement on their first judgments (Cohen's K ranging from .88 to .96, all $p<.001$). For the final score they discussed each item on which they disagreed, until reaching a full agreement. Two independent judges examined the gestures produced by each participant in the recollection phase and reached a significant level of agreement in recognizing 91% of the hand movements performed as gestures (Cohen's K ranging from .91 to .94, all $p<.001$). For the final score, the judges discussed each item on which they disagreed, until reaching a full agreement. For example, the following movement was considered to be an accompanying gesture: raising the right hand in front of the body, to about chest level, palm down, fingers joined at the tips; moving the hand and forearm rapidly, in a horizontal direction and by about 40 centimeters to the right. This movement was performed when saying the words "color has a very direct effect" (correct recollection).

The results for each type of recollection are shown in Table 3. In line with our expectations, in the recollection phase, participants in the Gesture condition produced fewer co-speech gestures (45% of recollections accompanied by gestures: 5.81 units accompanied by gestures out of 12.79 recollected units) than participants in the No Gesture condition (78% of recollections accompanied by gestures: 7.46 units accompanied by gestures out of 9.61 recollected units; t-test for independent samples: $t(28)=-7.73$, tied $p<.001$). More in detail, as regards correct recollections, participants in the Gesture condition produced gestures to accompany fewer recollected units than those in the No Gesture condition (t-test for

independent samples: t(28)=-7.74, tied p<.001). As regards errors, discourse-based inferences and elaborative inferences, there was no significant difference in the production of gestures by participants in the two experimental conditions (t-test for independent samples: t(28) ranging from 0 to 1.36, p ranging from .19 to 1).

As for Experiment 1, we investigated whether the semantic gestures and beats produced by the actor facilitated the number of proper recollections (i.e., correct recollections and discourse-based inferences) to the same extent. For that purpose we only considered those semantic units that the actor accompanied with a single type of gesture. We thus obtained 23 semantic units: 13 accompanied by semantic gestures (all representational gestures) and 10 by beats. The mean numbers of proper recollections according to the dichotomy beats-semantic gestures, are shown in Table 4. The results revealed that the actor's semantic gestures facilitated proper recollections significantly better than beats (paired t-test: t(14)=2.23, p=.02). A more detailed analysis of the semantic gestures also revealed that the number of proper recollections varied according to the gestures with which the actor had accompanied the respective semantic units (ANOVA: $F_{2,20}$ =4.21, p=.03); in particular, metaphoric gestures were associated with correct recollections more than iconic gestures (Scheffe: p= .03).

Lastly, we investigated whether the listeners properly recollected concepts that the speaker had accompanied with semantic gestures better than those accompanied by beats. In properly recalling the 10 semantic units that the actor had accompanied with one or more beats, the listeners produced a mean of 2.8 (sd=2.73) accompanying gestures, while in properly recalling the 13 semantic units that the actor had accompanied with one or more semantic gestures, they produced a mean of 1.09 (sd=1.28) accompanying gestures. As for Experiment 1, the results revealed that listeners produced more gestures when reporting concepts that the actor had accompanied with beats than when reporting concepts accompanied by semantic gestures (paired t-test, t(14)=-2.16, p=.04).

As for Experiment 1, we conducted an analysis in order to exclude the possibility of the influential factor being the verbal content uttered by the actor rather than the gestures he used. We examined the proper recollections of the 23 units by participants in the No Gesture condition, and compared the number of proper recollections of the 13 units in the Gesture condition accompanied by semantic gestures with the number of proper recollections of the 10 units in the Gesture condition accompanied by beats. In this case we found no difference between the number of proper recollections (paired t-test, $t(14)=1.50$, $p=1.36$).

Discussion and Conclusions

The analysis of the words and gestures produced by the participants in our experiments supports the hypothesis that spontaneous gestures by the speaker facilitate the construction of a rich and articulated model of the discourse, which in turn involves the use of fewer gestures by the listener in the recollection stage. Participants in the Gesture condition of the two experiments produced a significantly lower number of spontaneous gestures than participants in the No Gesture condition when recalling information correctly. The same result did not hold for discourse-based inferences, elaborative inferences and errors. A possible explanation is that these recollections were produced very infrequently, regardless of the experimental condition (Gesture or No Gesture); this may thus be the reason why we found no significant differences in the two conditions.

We also conducted a more detailed analysis to investigate whether the type of gesture produced by the actor influenced the quality of the mental model in relation to the concept accompanied by that gesture. The results revealed that semantic gestures facilitate proper recollections (that is, correct recollections and discourse-based inferences) more than beats, suggesting that semantic gestures contribute more to the construction of a complete mental model than beats. Moreover, we investigated whether listeners gesticulated differently in the recollection stage when recalling concepts that were originally accompanied by semantic

gestures or beats. The results revealed that listeners gesticulate more when recalling concepts originally accompanied by beats. This finding is consistent with the previous one in suggesting that the parts of the actor's discourse accompanied by semantic gestures are more effectively represented in the mental model of the discourse, and thus require a reduced cognitive load when expressed.

Taken together, these findings clarify the role of one factor that affects the production of gestures by speakers, namely the quality of the mental model of the content of the discourse. Nevertheless, other factors have an important role in determining the production of gestures: we know, for instance, that speakers tend to produce more co-speech gestures to convey non-verbal information when they are in an intermediate stage of the knowledge-acquisition process (Church & Goldin-Meadow, 1986) compared to when they are in the end-state of learning. This result may suggest that a learner’s gestures reflect his need to organize an articulated mental model of the material to be learnt. We also know that reciprocal visibility between the speaker and the listener affects gesture production: a speaker who can see the listener tends to produce more gestures than one who cannot (Mol et al., 2008). This result can be interpreted by assuming that the speaker voluntarily produces a large number of gestures in an attempt to facilitate the listener's comprehension.

We can conclude that the purpose of co-speech gestures, from the perspective of both the listener and the speaker, is to support the organization of thoughts through the construction of a mental model of the discourse. Our results, together with the findings by Cutica and Bucciarelli (2008), suggest that deep comprehension of a discourse (by the listener) and of the organization of the discourse (by the speaker) relies on the construction of complete and articulated mental models, made possible by integrating discrete information, namely linguistic information, and non-discrete information, namely the gestures accompanying a discourse. Future research should investigate the function of gestures in

adding semantic information to the discourse model in parallel with the function of gestures in directing the listener's attention to the verbal information they accompany.

Acknowledgements

The authors were supported in this research by Regione Piemonte of Italy, ATLAS project ID 44.

References

Alibali, M.W., & DiRusso, A.A. (1999). The function of gestures in learning to count: More than keeping track. *Cognitive Development, 14*, 37-56.

Alibali, M.W., Flevares, L., & Goldin-Meadow, S. (1997). Assessing knowledge conveyed in gesture: Do teachers have the upper hand*? Journal of Educational Psychology, 89,* 183-193.

Alibali, M.W., Kita, S., Bigelow, L.J., Wolfman, C.M., & Klein, S.M. (2001). Gesture plays a role in thinking for speaking. *Actes du Colloque, Orange*, (pp. 407-410). Aix-en-Provence, 18-22 juin.

Alibali, M.W., Kita, & S.,Young, A.J. (2000). Gesture and the process of speech production: We think, therefore we gesture. *Language & Cognitive Processes, 15*, 593-613.

Biggs, J.B. (2001). Enhancing learning: A matter of style or approach ? In: R.J. Sterneberg and L-F. Zhang (Eds.), *Perspectives on Thinking, Learning, and Cognitive Styles*, (pp. 73-102). N.J: LEA.

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

Bucciarelli, M. (2007). How the construction of mental models improves learning. *Mind & Society, issue 1, vol. 6,* 67-89.

Bull, P., & Connelly, G. (1985). Body movement and emphasis in speech. *Journal of Nonverbal Behavior*, 9, 169-187.

Butterworth, B., & Hadar, U. (1989). Gesture, speech, and computational stages: A reply to McNeill. *Psychological Review, 96*, 168-174.

Cassell, J. (1998). A Framework For Gesture Generation And Interpretation. In R. Cipolla, A. Pentland, (Eds.), *Computer Vision in Human-Machine Interaction*, (pp. 191-215). New York: Cambridge University Press.

Church, R.B., & Goldin-Meadow, S. (1986). The mismatch between gesture and speech as an index of transitional knowledge. *Cognition, 23*, 43-71.

Cook, S.W., Jaeger, T.F., & Tanenhaus, M.K. (2009). Producing Less Preferred Structures: More Gestures, Less Fluency. *The 31st Annual Meeting of the Cognitive Science Society*, (pp. 62-67). Amsterdam, July, 2009.

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

Efron, D. ([1941] 1972). *Gesture, Race and Culture*. The Hague: Mouton.

Ekman, P., & Friesen, W.V. (1972). Hand movements. *Journal of Communication, 22*, 353-374.

Feyereisen, P., & Harvard, I. (1999). Mental imagery and production of hand gestures while speaking in younger and older adults. *Journal of Nonverbal Behavior*, *23*, 153-171.

Feyereisen, P., Van de Wiele, M., & Dubois, F. (1988). The meaning of gestures: What can be understood without speech. *Cahiers de Psychologie Cognitive, 8*, 1, 3-25.

Garnham, A., Oakhill J., & Cain K., (1998). Selective retention of information about the superficial form of text: ellipses with antecedents in main and subordinate clauses. *The Quarterly Journal of Experimental Psychology: Section A, 51(1)*, 19-39.

Gentilucci, M., & Dalla Volta, R. (2008). Spoken language and arm gestures are controlled by the same motor control system. *The Quarterly Journal of Experimental Psychology, 61*, 944-957.

Gyselinck, V., & Tardieu, H. (1999). The role of illustrations in text comprehension: What, when, for whom and why? In H. Van Oostendorp, S. Goldman (Eds.), *The construction of mental representations during reading*, (pp. 195-218). Mahwah, NJ: Erlbaum Associates.

Goldin-Meadow, S. (1999). The role of gesture in communication and thinking. *Trends in*

Cognitive Science, 3(11), 419-429.

Goldin-Meadow S., & Alibali, M.W. (1999). Does the hand reflect implicit knowledge? *Behavioral and Brain Sciences, 22(5),* 766-787.

Goldin-Meadow, S., Wein, D., & Chang, C. (1992). Assessing knowledge through gesture: Using children's hands to read their minds. *Cognition and Instruction, 9* (3), 201-219.

Goodwin, C. (1986). Gesture as a resource for the organization of mutual orientation. *Semiotica, 62*(1-2):29-49.

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.

Hadar,U., Burnstein, A. Krauss, R.M., & Soroker, N. (1998). Ideational gestures and speech: A linguistic investigation. *Language and Cognitive Processes, 13*, 59-76.

Hildebrandt, B., Moratz, R., Rickheit, G., & Sagerer, G. (1999). Cognitive modelling of vision and speech understanding. In G. Rickheit and C. Habel (Eds.), *Mental models in discourse processing and reasoning*, (pp. 213-236). Elsevier, North Holland.

Iverson, J. M., & Goldin-Meadow, S. (1998). Why people gesture when they speak. *Nature, 396*, 228.

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*. Lawrence Erlbaum Associates Ltd, London.

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

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., & Church, R. B. (1997). Can children detect conceptual information conveyed through other children's nonverbal behaviors? *Cognition and Instruction,* 15(1), 107-134.

Kelly, S., & Church, R.B. (1998). A comparison between children's and adults' ability to detect conceptual information conveyed through representational gestures. *Child Development, 69*, 85-93.

Kendon, A. (1983). Gesture and speech: How they interact. In J. A. Wiemann, R.P. Harrison (Eds), *Nonverbal Interaction*, 13-45. Beverly Hills, London: Sage Publications.

Kendon, A. (1994). Do gestures communicate? A review. *Research on Language and Social Interaction, 27*, 175-200.

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

Kita, J.S. (Ed.) (2000). *Pointing: where language, culture and cognition meet*. New Jersey, Lawrence Erlbaum Associates.

Kita, J. S., & Lausberg, H. (2008). Speech-gesture discoordination in split brain patients' left-hand gestures: Evidence for right-hemispheric generation of co-speech gestures. *Cortex, 44*, 131-139.

Krauss, R. M. (1998). Why do we gesture when we speak? *Current Directions in Psychological Science, 7*, 54–60.

Krauss, R.M., Chen, Y., & Chawla, P. (1996). Nonverbal behavior and nonverbal communication; What do conversational hand gestures tell us? In M. Zanna, (Ed.), *Advances in experimental social psychology,* (pp. 389-450). San Diego, CA: Academic Press.

Krauss, R. M., Chen, Y., & Gottesman, R. (2000). Lexical gestures and lexical access: A

process model. In D. McNeill (Ed.), *Language and gesture*, (pp. 261-283). Cambridge: Cambridge University Press.

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. (1987). *Psycholinguistics: A new approach*. New York: Harper & Row.

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

McNeill, D., Cassell, J., & McCullough, K.E. (1994). Communicative effects of speech mismatched gestures. *Research on Language and Social Interaction, 27*, 223-237.

Mol L., Maes A., Krahmer E., & Swerts M., (2008). Cognitive Effort and Gesticulation. *Speech and Face to Face Communication. A workshop/summer school dedicated to the memory of Christian Benoît*, 27-29 October 2008, Grenoble, France.

Rauscher, F. H., Krauss, R. M., & Chen, Y. (1996). Gesture, speech, and lexical access: The role of lexical movements in speech production. *Psychological Science, 7*, 226–231.

Rickheit, G., & Sichelschmidt, S. (1999). Mental models: some answers, some questions, some suggestions. In G. Rickheit, C. Habel (Eds.), *Mental models in discourse processing and reasoning*, (pp. 9-40), North-Holland, Elsevier Science, Amsterdam.

Rimè B. (1982). The elimination of visible behavior from social interactions: effects of verbal, nonverbal and interpersonal variables. *European Journal of Social Psychology, 12*, 113-129.

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

Thompson, L.A, Driscoll, D., & Markson, L. (1998). *Journal of Nonverbal Behavior, 22*, 167-187.

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

Academic 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. Experiment 1: Means of recollections accompanied by gestures out of the global recollections and according to the type of recollection*.

Condition	Correct recollections	Discourse-based inferences	Elaborative inferences	Errors
Gesture (N=19)	9.05/21.89	.53/1.05	.21/.95	.26/.53
	(41%)	(50%)	(4%)	(49%)
No Gesture (N=19)	8.68/15.32	.16/.21	.16/.58	.32/.68
	(58%)	(76%)	(27%)	(47%)

* (in parentheses the percentages of recollections accompanied by gestures)

Table 2. Experiment 1: Means of proper recollections (correct recollections and discourse-based inferences) by participants for each actor's gesture category (standard deviation in parenthesis).

Actor's gestures	**Participants' proper recollections**		
Beats (n=12)	6.31 (3.95)		
Semantic gestures (n=21)	8.78 (3.75)	Metaphoric (n=7)	11.99 (3.65)
		Deictic (n=5)	7.91 (3.87)
		Iconic (n=9)	6.51 (2.94)

Table 3. Experiment 2: Means of recollections accompanied by gestures out of the global recollections and according to the type of recollection*.

Condition	Correct recollections	Discourse-based inferences	Elaborative inferences	Errors
Gesture (N=15)	4.87/10.73	.47/.93	.20/.60	.27/.53
	(45%)	(50%)	(33%)	(50%)
No Gesture (N=15)	6.80/7.60	.20/.27	.13/.87	.33/.87
	(88%)	(74%)	(14%)	(38%)

* (in parentheses the percentages of recollections accompanied by gestures)

Table 4. Experiment 2: Means of proper recollections (correct recollections and discourse-based inferences) by participants for each actor's gesture category (standard deviation in parenthesis).

Actor's gesture	**Participants' correct recollections**		
Beats (N=10)	2.97 (1.34)		
Semantic gestures (N=13)	4.53 (2.05)	Metaphoric (N=8)	6.28 (2.74)
		Iconic (N=5)	2.30 (1.23)

Appendix 1 : Excerpt from the discourse in Experiment 1

(Semantic units are separated by slashes)

It was there, at the funfair, it was there that I found her, and it was at the funfair that I lost her. It was a vast funfair./A funfair with shooting-ranges and candy floss stalls and Japanese bagatelle tables, stalls with bottles of champagne and showmen's booths and roundabouts./And the roundabouts turned and creaked and the candy floss scented the air and the rifles shot./I was shooting at the target./I can shoot at the target very well and I am proud of it./No, wait a moment, I am wrong! I did not meet her at the shooting-range./I met her at the candy floss stall. Yes, it was at the candy floss stall that I found her./The candy floss scented the air,/and she was eating it and she blew on her candy and I was all covered with white powder./She started laughing/and I asked her: "What's your name?"/And she shouted to me: "I'll tell you later"./Later, we went to the shooting-range/and it was there that I lost her./I aimed at the target, breaking all the clay pipes/and every time she shouted to me: "Well done!"/And then, when there were no pipes left to break, I aimed at the egg, the one held up by the jet of water,/and while I was aiming I shouted to her: "What's your name?"/And she replied: "I'll tell you later."/I shot, the egg popped up./I turned aside and she wasn't there any more.

Appendix 2: Excerpt from the discourse in Experiment 2

(Semantic units are separated by slashes)

It's beyond dispute that colors carry strong expressive components./Some attempts have been made to describe the specific expressive characters of the various colors and to draw some general conclusions from the symbolic use the different cultures have made of them./There is a very widespread belief that the expression of colors is based on association./Therefore, red should be considered exciting because it reminds us of the connotations of fire, blood and revolution./Green evokes the restorative thought of nature,/and blue is refreshing like water./However, the theory of association is not more interesting or prolific in this field than in others. The effects of colors are too direct and spontaneous to be simply the results of a interpretation given through knowledge./On the other hand, no hypothesis has been advanced so far on the kind of physiologic process which could help to explain the influence of colors on the organism./ It is well known that extreme brightness, high saturation and shade of color corresponding to vibrations in wavelength breadth cause excitement./

Appendix 1 : Excerpt from the discourse in Experiment 1

(Semantic units are separated by slashes)

It was there, at the funfair, it was there that I found her, and it was at the funfair that I lost her. It was a vast funfair./A funfair with shooting-ranges and candy floss stalls and Japanese bagatelle tables, stalls with bottles of champagne and showmen's booths and roundabouts./And the roundabouts turned and creaked and the candy floss scented the air and the rifles shot./I was shooting at the target./I can shoot at the target very well and I am proud of it./No, wait a moment, I am wrong! I did not meet her at the shooting-range./I met her at the candy floss stall. Yes, it was at the candy floss stall that I found her./The candy floss scented the air,/and she was eating it and she blew on her candy and I was all covered with white powder./She started laughing/and I asked her: "What's your name?"/And she shouted to me: "I'll tell you later"./Later, we went to the shooting-range/and it was there that I lost her./I aimed at the target, breaking all the clay pipes/and every time she shouted to me: "Well done!"/And then, when there were no pipes left to break, I aimed at the egg, the one held up by the jet of water,/and while I was aiming I shouted to her: "What's your name?"/And she replied: "I'll tell you later."/I shot, the egg popped up./I turned aside and she wasn't there any more.

Appendix 2: Excerpt from the discourse in Experiment 2

(Semantic units are separated by slashes)

It's beyond dispute that colors carry strong expressive components./Some attempts have been made to describe the specific expressive characters of the various colors and to draw some general conclusions from the symbolic use the different cultures have made of them./There is a very widespread belief that the expression of colors is based on association./Therefore, red should be considered exciting because it reminds us of the connotations of fire, blood and revolution./Green evokes the restorative thought of nature,/and blue is refreshing like water./However, the theory of association is not more interesting or prolific in this field than in others. The effects of colors are too direct and spontaneous to be simply the results of a interpretation given through knowledge./On the other hand, no hypothesis has been advanced so far on the kind of physiologic process which could help to explain the influence of colors on the organism./ It is well known that extreme brightness, high saturation and shade of color corresponding to vibrations in wavelength breadth cause excitement./