

Non-determinism in the Uptake of Gestural Information

Ilaria Cutica^a, Monica Bucciarelli^b

^aDepartment of Health Sciences, University of Milan

Via Antonio di Rudini, 8, – 20140, Milan, Italy

e-mail: ilaria.cutica@unimi.it

^bDepartment of Psychology and Center for Cognitive Science, University of Turin

Via Po, 14 –10123 Turin, Italy

e-mail: monica.bucciarelli@unito.it

Corresponding author:

Ilaria Cutica

University of Milan, Department of Health Sciences

Via Antonio di Rudini 8 - 20140 Milan, Italy

e-mail: ilaria.cutica@unimi.it

tel: +39.02.50318938; fax:+39.02.91533899

Disclosure Statement. The authors declare they do not have personal nor institutional conflicts of interest.

Abstract

It is well established that gestures and speech form an integrated system of communication; gestures that match the meaning of the speech they accompany favor the listener's discourse comprehension, whereas mismatching gestures whose meaning conveys information contradicting that conveyed by speech, impair comprehension. A less investigated issue is whether or not the uptake of gestural information is a deterministic process. In line with recent studies in the literature, we purport that the process may be modulated by certain factors. In particular, we investigate the role of unrelated gestures whose meaning, which is irrelevant to the speech they accompany, could be neglected. The results of four experiments led us to conclude that unrelated gestures are not processed, and that the uptake of gestural information is a non-deterministic process.

Keywords: co-speech gestures; discourse comprehension; mental models; unrelated gestures; multimodal information processing

Introduction

When people talk, they move their hands and arms; these movements, which are not conventionalized, are called co-speech gestures. Starting from the seminal works of McNeill (1985, 1992, 2005) and Kendon (2004), it is generally agreed that co-speech gestures and speech function together as a unit, by forming a single, integrated system of communication (see also Church, Kelly & Lynch, 2000; Goldin-Meadow, Kim & Singer, 1999; McNeil, Alibali & Evans, 2000; McNeill, Cassell & McCullough, 1994; but cf. Hadar, Wenkert-Olenik, Krauss & Soroker, 1998; Krauss & Hadar, 1999). Indeed, the literature reveals that co-speech gestures have a function for both the speaker and the listener (for a review see Goldin-Meadow & Alibali, 2013).

From the listener's viewpoint, the speaker's co-speech gestures are functional to deep comprehension and learning: several studies have revealed that discourse comprehension benefits from such gestures (see for instance Alibali, Flevares, & Goldin-Meadow, 1997; Iverson & Goldin-Meadow, 2001; Kelly, Barr, Church & Lynch, 1999; Kelly & Church, 1998; McNeil, Alibali & Evans, 2000). Speakers' gestures facilitate speech-reading, namely the perception of what a person says by observing the movements of her/his lips (Berger & Popelka, 1971; Vendrame, Cutica & Bucciarelli, 2010). Further, co-speech gestures facilitate memory for sentences (Thompson, Driscoll & Markson, 1998), and listeners' comprehension of degraded (Riseborough, 1981), ambiguous (Thompson & Massaro, 1986) or highly complex (McNeil et al., 2000) verbal messages (see also Wu & Coulson, 2010). A recent quantitative meta-analysis that included 63 separate samples found that gestures foster comprehension in listeners (Hostetter, 2011), although the size of the beneficial effect depends on several factors, including the topic of the gestures, their semantic overlap with speech, and the age of the listeners.

It is well established that listeners integrate information from gestures into their semantic interpretation of speech (see e.g., Alibali et al., 1997; Beattie & Shovelton, 1999, Goldin-Meadow, 1999; Kelly et al., 1999; McNeill et al., 1994). Indeed, when the meaning of gestures adds, specifies, or reinforces information conveyed by speech, gestures facilitate speech comprehension, but when gestures convey information contradicting that conveyed by speech, they hinder the comprehension of that speech (Cassell, McNeill & McCullough, 1999; Goldin-Meadow et al., 1999). Studies on the neural correlates of co-speech gestures have also revealed that co-speech gestures affect semantic processing of the accompanied words (e.g., Wang & Chu, 2013), and that, conversely, the semantic content of speech is relevant in co-speech gesture processing (see Willems & Hagoort, 2007 for a review). Such studies have also shown that the brain produces different responses (i.e., a larger N400, a negative deflection between 200 and 500 msec from the stimulus onset, considered as a marker of semantic integration) when gestures convey the same information as speech compared to when they convey different information (e.g., Kelly, Kravitz & Hopkins, 2004), and when speech is accompanied by self-grooming movements with respect to no hand movements or meaningful speech-associated gestures (Skipper, Goldin-Meadow, Nusbaum & Small, 2007). Furthermore, speech and gesture are integrated most efficiently when the differences in onsets do not exceed a certain time span, as the co-occurrence of gestures and speech is useful to disambiguate gestures (Habets, Kita, Shao, Özyurek & Hagoort, 2011).

The findings in the literature have led to the assumption that listeners cannot help but integrate gestural information (see, e.g., Cassell et al., 1999). However, the need to test the assumption that the uptake of gestural information is deterministic and unavoidable has only recently been addressed. In favor of a non-deterministic process, the results of some studies have suggested that addressees' uptake of gestural information is modulated by speakers' gaze (Gullberg & Kita, 2009) and intentional relationship between gesture and speech (Kelly,

Ward, Creigh & Bartolotti, 2007). The results of ERP studies have also suggested that the integration of gesture and speech in comprehension is not an automatic and obligatory process, but is modulated by situational factors (Holle & Gunter, 2007).

In this paper, we investigate whether the listener can disregard the information conveyed through gestures when these are unrelated to the accompanying speech, namely when their meaning is unrelated to the spoken words. We present a mental model account of the role of gestures in discourse comprehension, relevant to establish what counts as a measure of discourse comprehension. We then explain why it is important to study the role of unrelated gestures in discourse comprehension in order to discover whether the uptake of gestural information is a deterministic process. Finally, we describe four experiments designed to establish whether or not the processing of gestural information is deterministic, and we draw the relevant conclusions.

A Mental Model Framework for Discourse Comprehension and Co-speech Gestures

Many researchers have argued that the successful comprehension of a discourse is tantamount to the construction of a coherent mental model (e. g., Garnham & Oakhill, 1992; Glenberg, Kruley & Langston, 1994; Graesser, Millis & Zwaan, 1997; Graesser, Singer, & Trabasso, 1994; Zwaan, Magliano & Graesser, 1995; Zwaan & Radvansky, 1998). According to different theoretical frameworks, such representations are referred to as the “mental model” (Johnson-Laird, 1983, 2006; Johnson-Laird & Byrne, 1991) or “situation model” (van Dijk & Kintsch, 1983; Kintsch, 1998). A mental model is an internal representation of a real or fictional state of affairs, and is usually built on-the-spot to deeply understand and reason about the state of affairs (Bucciarelli & Cutica, 2012). Mental models are special sorts of representations that, according to Johnson-Laird (1983, 2006), have three characteristics distinguishing them from other mental representations (e.g., semantic networks, expressions in a language of thought or predicate calculus, connectionist webs of associations):

1. Each mental model represents a set of possibilities that have in common what the model represents. Thus, as compared to mental images, models are more abstract representations.

2. Mental models are iconic, i.e., their structure corresponds to the structure of what they represent.

3. Mental models are parsimonious; they represent what is true, possible, according to the premises, but by default not what is false, impossible.

As Cutica and Bucciarelli (2012) point out, mental models have been invoked as an explanatory principle for comprehension processes at a text/discourse level. Different depths of processing of a discourse can be detected on a continuum, starting from shallow levels - at which the listener constructs a mental representation of the discourse in a propositional format - up to deep levels of processing - at which the listener creates an articulated mental model of the contents of the discourse. A discourse mental model represents the state of affairs to which a discourse refers by integrating temporal, spatial, causal, motivational, person, and object-related information stated explicitly in the discourse. Thus, a discourse mental model captures the discourse significance.

Cutica and Bucciarelli (2008; see also Bucciarelli, 2007) argued that co-speech gestures favor the construction of the discourse model because the information they convey is easily incorporated into that model. Indeed, as the information conveyed by co-speech gestures is represented in a non-discrete format, it is particularly suitable to be integrated into a model representation because mental models themselves are non-discrete iconic representations of a certain situation (see Hildebrandt, Moratz, Rickheit & Sagerer, 1999; Rickheit & Sicheleschmidt, 1999). Non-discrete representations, unlike proposition-like representations, are non-linguistic and analogical representations whose structure resembles the structure of the states of affairs represented (Bonatti, 1998).

The discourse mental model, enriched by gestural information, leads to better comprehension and retention of the discourse content at the expense of verbatim memory. This is due to the fact that mental models encode little or nothing of the linguistic form of the sentences on which they are based (see Johnson-Laird, 1983; Zwaan & Radvansky, 1998). Consistently, several experimental findings have revealed that individuals who build an articulated mental model of a given text have poorer verbatim recall of material than individuals who build a less articulated mental model (see Garnham, Oakhill, & Cain, 1998; Johnson-Laird & Stevenson, 1970).

According to the literature, the quality of a mental model may be measured in terms of correct information recalled and correct inferences drawn from the information explicitly contained in the discourse (Johnson-Laird, 1983, 2006; Johnson-Laird & Byrne, 1991). More in detail, the inferences that may be considered as indicators of a rich mental model are discourse-based inferences, which make explicit that information which is originally implicit in the text and 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 et al., 1994). Discourse-based inferences differ from elaborative inferences (e.g., Singer, 1994), that are instead a sort of personal enrichment of the original text, and are not considered to be indicators of the quality of mental models.

Consistent with the assumptions of the mental model theory on the role of gestures in discourse comprehension, Cutica and Bucciarelli (2008) found that participants listening to a discourse accompanied by gestures recalled more correct information, drew more discourse-based inferences, and recognized the literal form of sentences less well than participants listening to the same discourse not accompanied by gestures.

Non-determinism in the Uptake of Gestural Information: Unrelated Gestures as a Test Base

According to the literature, gestures are however processed together with the speech they accompany, even though the meaning they convey is inconsistent with speech. Several studies (e.g., Alibali et al., 1997; Cassell et al., 1999; McNeill et al., 1994; Kelly & Church, 1998; Thompson & Massaro, 1994) have shown that gestures whose meaning contrast with the meaning conveyed through speech, impair discourse comprehension because they result in the listener trying to reconcile the two contrasting meanings (McNeill et al., 1994; see also Cassell et al., 1999). Other studies reveal that seeing incongruent gestures may even lead to a better gist recall, albeit accompanied by the reproduction of the incongruent gesture, and by inaccuracies in speech (e.g., Galati & Samuel, 2011). The attempt to reconcile the incongruent meanings of speech and gesture involves processing costs, as shown by ERP studies (Kelly et al., 2004; Kelly et al., 2007; Ozyurek et al., 2007; Wu & Coulson, 2005).

In our view, to investigate whether gestures are always processed together with speech, that is to reveal the possible non-deterministic nature of the uptake of gestural information, it is necessary to use gestures that do not have any sort of semantic connection with the speech they co-occur. On the contrary, incongruent gestures, as described in the literature relevant to our investigation, maintain a sort of semantic connection with the speech they accompany, and for this reason are not apt to reveal whether gestures are obligatorily processed together with speech. This claim holds for all the studies in the literature that, in our view, are concerned with three types of incongruent gestures: 1) mismatching gestures conveying information differing from that conveyed by the accompanying speech, but that is potentially integrable with that information. Consider, for example, calling the water in a container ‘tall’ while indicating its width in gesture. The term ‘mismatch’ was originally introduced by Church and Goldin-Meadow (1986) for this kind of gestures, which maintain a

connection with the content conveyed by the speech and can therefore affect the information that is gleaned from speech (e.g., Goldin-Meadow & Sandhofer, 1999), although it is not clear whether they incur processing costs. Since then, the term ‘mismatch’ has been used to refer to different kinds of speech-gestures combinations that, in our view, still maintain a semantic connection with the speech they accompany: mismatching gestures that contradict the information conveyed by speech, and mismatching gestures that introduce referential ambiguity. 2) Mismatching gestures that contradict the information conveyed by speech refer to the same conceptual domain as speech. Consider, for example, the case of a mismatch in which the speaker says “up” and “above” while producing gestures depicting respectively “down” and “below” (McNeil et al., 2000, p. 138), or the case in which the speaker says “she offers him a penny” while producing a gesture depicting her giving something to herself (Cassell et al., 1999). The speech and the co-occurring gesture cover the same conceptual domain (i.e., in our examples, direction and addressee), although they convey a contrasting content: their meaning is not unrelated to the spoken words. In other words, this kind of mismatching gestures are however salient with respect to the conceptual domain of the idea expressed in speech, and their meaning may be processed together with the meaning of the co-occurring speech. 3) Mismatching gestures that introduce referential ambiguity also maintain a sort of connection with the corresponding speech. As a case example, consider the gestures involved in the ERP study conducted by Kelly et al. (2007). The participants in the experiment watched videos of two different people describing two objects verbally and by gesticulating: a tall, thin glass and a short, wide dish. In the congruent condition, the speech and gesture communicated the same information about the same object (e.g., saying tall and gesturing to the tallness of the glass). In the incongruent condition, the speech and gesture presented different information about two different objects (e.g., saying tall and gesturing to the shortness of the dish). Mismatching gestures thus gave rise to ambiguity as regards the

actual conceptual unit: the speech referred to one object, and the gesture to a different one; however, both objects were possible referents.

In our view, the deterministic versus non-deterministic nature of the uptake of gestural information can be tested by conducting studies using gestures that are fully arbitrary with respect to the co-occurring speech; this means that they refer to a different conceptual domain than the speech, and that they do not introduce a referential ambiguity. We call such gestures *unrelated*, because their meaning is unrelated to the spoken words. If unrelated gestures have a similar effect on listeners to that of mismatching gestures, then we can conclude that gestures are obligatorily processed; on the contrary, if unrelated gestures do not impair comprehension then we can conclude that they are not semantically processed.

To our knowledge, only one study by Feyereisen (2006) has investigated what we call unrelated gestures, albeit on memory for sentences and under the label of ‘mismatching gestures’. For the purpose of our study it was instead important to manipulate the ease with which the participants in the experiment were able to build an articulated mental model of a full discourse. Hence, in our study we investigated memory for discourse.

Since it is impossible to ask a person to speak while producing unrelated gestures, we adopted the following procedure. We created a computer animated (male) agent who produced a discourse in three different conditions. In each condition, his co-speech gestures either: (a) matched the speech (*gesture condition*); or (b) were unrelated to the speech (*unrelated condition*); or (c) were absent (*no gesture condition*).

The deterministic and the non-deterministic hypotheses lead to different predictions in recognition memory tasks. By the non-deterministic hypothesis, performance in the unrelated and no gesture conditions should be comparable, because in these conditions participants do not process any gestures. By the same hypothesis, participants should perform better in recognizing discourse verbatim in the unrelated and no gesture conditions as

compared to the gesture condition, because in the latter condition they construct an articulated mental model which implies losing verbatim memory for discourse. On the contrary, by the deterministic hypothesis, participants should perform better in recognizing discourse verbatim in the no gesture condition as compared to the gesture condition, because they constructed a less articulated discourse mental model, but should perform less well in the unrelated condition, because unrelated gestures, if processed, add difficulty to discourse processing.

In recall memory tasks, the non-deterministic hypothesis and the deterministic hypothesis lead to different predictions about accuracy in terms of correct recollections and discourse-based inferences. By the non-deterministic hypothesis, participants should perform better in the gesture condition than in both the unrelated and the no gesture conditions, and should demonstrate comparable levels of performance in the unrelated and no gesture conditions. On the contrary, by the deterministic hypothesis, participants should perform less well in the unrelated condition than in the no gesture condition because unrelated gestures, if processed, add difficulty to discourse processing. Furthermore, participants should perform best in the gesture condition.

Experiment 1: Gestures in Memory for Discourse: Recognition Memory

If the uptake of gesture information is non-deterministic, then a discourse accompanied by unrelated gestures (unrelated condition) or delivered without gesturing (no gesture condition), as compared with a discourse produced with matching gestures (gesture condition), results in better performance in terms of verbatim recognition of sentences in the discourse.

Participants

Sixty students from Turin University voluntarily took part in the experiment (38 females and 22 males, mean age: 22), twenty were randomly assigned to the Gesture condition, twenty to the Unrelated condition and twenty to the No Gesture condition.

Materials and Procedures

The experimental material comprised a videotaped discourse presented in Italian by an animated agent. The discourse was part of one used by Cutica and Bucciarelli (2008): it was a narrative discourse concerned with a series of events that occurred at a funfair (see Appendix 1A), and lasted about two minutes. The discourse was presented in three different conditions, for a total of three videotaped fictions. In the gesture condition the animated agent accompanied the discourse with co-speech gestures (movements of one or both hands, arms, and shoulders); the gestures performed by the animated agent were modeled as an accurate copy of the gestures performed by the human actor in the original experiment by Cutica and Bucciarelli (2008), who had been instructed to produce hand and arm movements as he felt appropriate with respect to the discourse flow. Such procedure was meant to ensure that we investigated naturally occurring gestures (see also Gullberg & Kita, 2009). Two judges examined the gestures produced by the agent and found that he produced 3 deictic gestures, 9 representational gestures, and 3 baton gestures¹; they also found that the agent never produced symbolic gestures (i.e., gestures with a widely recognized conventionalized meaning). However, at our level of analysis, we did not distinguish the effect of these three different types of gestures. Finally, the judges excluded the possibility that any of the gestures conveyed information that was implicit or absent in the co-occurring speech. This procedure

¹ There is a general agreement in distinguishing three main categories of co-speech gestures, albeit under different labels (see, e.g., Bangerter, 2004; Ekman & Friesen, 1972; Hadar, Burnstein, Krauss, & Soroker, 1998; Kendon, 1983; Krauss, Chen & Gottesman, 2000; McNeill et al., 1994): (1) *Deictic gestures*, which locate some aspects of the story being narrated in the physical space in front of the narrator, and establish a joint focus of attention with the addressee; (2) *Representational gestures* which pictorially represent concrete images of the speaker's thoughts (*iconic gestures*), or pictorially represent an abstract concept (*metaphoric gestures*); (3) *Batons* which refer to the rhythm of speech, and tend to have the same form regardless of the content.

was necessary to ascertain that the actor's gestures were not the source of any of the participants' discourse-based inferences. In the unrelated condition the animated agent's gestures appeared in the same sequence as the gestures in the gesture condition, but with a delay of about 20 seconds with respect to the corresponding speech: the gesture track simply moved 20 seconds forward regardless of what was going on in speech. The last gestures in the gesture condition were paired with the beginning of the discourse in the unrelated condition. We chose a delay of 20 seconds in order to disrupt the correspondence between the spoken content and the corresponding gesture(s) and therefore exclude the possibility of the listener analyzing the verbal and relative visual information (namely speech and co-speech gestures) together. The underlying assumption is that the duration of working memory does not exceed 20 seconds (see, e.g., Brown, 1958; Peterson & Peterson, 1959). As a side effect, the 20-second speech-gesture delay guarantees that in the unrelated condition gestures do not maintain synchrony with discourse in ways that reflect topic-focus structures or stress patterns. In the gesture and unrelated conditions, overall considered, the number and type of gestures were the same.

Two judges examined each gesture-speech combination produced by the actor in the unrelated condition, and they evaluated the gesture-speech combinations according to the gesture-speech segmentation (an excerpt of the segmentation is reported in Appendixes 1A and 1B). The two judges were asked to classify each gesture-speech combination as mismatching (gestures that conveyed information different from that conveyed by speech but still potentially integratable with that information) or unrelated. They both classified all the gesture-speech combinations as unrelated, and none as mismatching. This procedure was necessary to ascertain that the actor's gestures were not perceived by the participants as mismatching rather than unrelated.

Finally, in the no gesture condition the animated agent delivered the discourse without gesticulating. His position was modeled on the posture of the human actor in the no gesture condition of the original study.

In each condition the animated agent maintained the same facial expressions and the same labial movements. His voice was the recorded voice of the human actor in the original experiments, and from the beginning and throughout the discourse, the labial movements of the animated agent matched his speech. Figure 1 shows three frames from the videotaped fictions used in the three experimental conditions. The first and second frames depict the gesture produced by the agent while saying the word “big” in the sentence “It was a very big funfair” in the gesture and unrelated conditions, respectively. The third frame refers to the actor in the no gesture condition while saying the same word. Examples of gestures produced by the actor are in Appendix 1B.

Participants, randomly assigned to one experimental condition, were invited to watch one of the videotaped fictions, after which they were presented with a list of sentences, one by one in random order. The participants were invited to consider whether or not the sentence was identical to the sentence actually spoken by the actor. The sentences presented, as compared with those in the discourse, were of the following sorts: (1) identical (*literally correct*); (2) with the same meaning, but said with different words (*paraphrases*); (3) inconsistent in meaning but said with almost the same words (*wrong content*). We created 24 sentences, with eight in each category (examples are included in Appendix 1C). The non-deterministic hypothesis predicts that participants in both the unrelated and no gesture conditions would perform better in terms of recognizing the sentences actually proffered by the actor, as compared with participants in the gesture condition. The hypothesis does not make predictions for the ability to refute paraphrases; having constructed an articulated mental model leads to formulate one’s own paraphrases, which do not necessarily correspond

to the paraphrases we created for the experiment. Also, the non-deterministic hypothesis makes no predictions for the ability to refute sentences inconsistent in meaning with the sentences in the discourse; the construction of an articulated mental model does not prevent possible misunderstandings. However, the non-deterministic hypothesis also predicts comparable performances with the three sorts of sentences in case of a discourse accompanied by unrelated gestures and a discourse delivered without gesturing.

We coded responses of “*Yes*” to literally-correct sentences, and responses of “*No*” to paraphrases and wrong content sentences as correct.

Results

Several analyses of variance using the Q of Cochran test revealed that the eight stimuli constituting each sentence category were comparable in terms of difficulty ($6.49 < Q < 12.63$; $.08 < p < .91$). Table 1 illustrates the mean correct responses in the three conditions.

(Table 1 about here)

We performed a one-way ANOVA that revealed a main effect of the experimental condition on the ability to correctly identify sentences at verbatim level ($F(2, 57) = 10.04$, $p < .0001$). On the contrary, the effect was not significant for the ability to refute paraphrases ($p = .06$) and wrong content sentences ($p = .12$). In line with the non-deterministic hypothesis, both the participants in the unrelated and the no gesture conditions performed better than the participants in the gesture condition in identifying sentences verbatim (Tukey post hoc: $p < .0001$, and $p = .02$, respectively). Further, the participants in the unrelated condition and in the no gesture condition showed comparable levels of performance in identifying literally correct sentences, and refuting paraphrases and wrong content sentences (Tukey post hoc: $.27 < p < .98$).

Experiment 2: Gestures in Memory for Discourse: Recollection Memory

If the uptake of gesture information is non-deterministic, then a discourse accompanied by unrelated gestures (unrelated condition) or delivered without gesturing (no gesture condition), as compared with a discourse produced with matching gestures (gesture condition), results in worse performance in terms of correct recollections and discourse based inferences. However, performance by participants in the unrelated condition should be comparable with performance by those in the no gesture condition. The non-deterministic hypothesis makes no predictions as regards the number of elaborative inferences and erroneous recollections, because mental models do not necessarily prevent a person from making mistakes: 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.

Participants

Sixty students from Turin University voluntarily took part in the experiment (40 females and 20 males, mean age: 22). Twenty participants were randomly assigned to the gesture condition, twenty to the unrelated condition and twenty to the no gesture condition. None of the participants in Experiment 2 had taken part in Experiment 1.

Materials and Procedures

The experimental material was the same as for Experiment 1, namely the videotaped discourse in the gesture, the unrelated, and the no gesture conditions. The experimental procedure was different. The participants were randomly assigned to one of the three experimental conditions; they were invited to watch the fiction, and were given the following instructions: “This is an experiment to investigate how people comprehend a discourse. I’m going to show you a video of a discourse by an animated agent. You must watch it carefully, paying attention to what he says. The discourse lasts approximately two minutes. At the end of the tape, I will ask you to tell me all you can remember of the discourse. Your answer will

be videotaped”. As soon as the video finished, the participants were asked to recall all they could remember (free-recall task); all of them were video-recorded.

To code the results, the discourse was divided into 21 semantic units, corresponding to as many main concepts that the listener could recall (see Appendix 1A). Each concept (i.e., semantic unit) recalled by the participants was evaluated according to the following coding schema:

- *Correct recollection*: a semantic unit recollected as a paraphrase².
- *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 recollection, consider, for instance, the following examples: “I met her at the candy floss stall. Yes, it was at the candy floss stall that I found her”. According to the coding schema, the statement “The man saw the girl for the first time at the candy floss” is a correct recollection, the statement “The man saw the girl for the first time at the shooting-range” is an erroneous recollection. The statement “It was such a vast funfair that the boy kept losing the girl” is a discourse-based inference with respect to the target-sentence: “It was a vast funfair”. Finally, the sentence “The boy said that he was shooting at the target, the round one that you have to hit in the middle, ...” is an elaborative inference with respect to the target-sentence: “I was shooting at the target”.

Results

² Although a correct recollection may assume the form of a literal recollection, we consider as correct recollections revealing the construction of the discourse mental model only those in which participants reformulate through their own words the content in the semantic units. Verbatim recalls tend to occur in the absence of such a model (see also the results of Experiment 1).

Two independent judges, blind with respect to the condition they were examining, coded the participants' recollections individually; 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 ($.90 < K < .92$, p always $< .001$). For the final score the judges discussed each item on which they disagreed, until reaching full agreement. Table 2 shows the mean scores for types of recollection in the three experimental conditions.

(Table 2 about here)

As a general result, and in all conditions, scores for correct recollections were higher than for other sorts of recollection. We performed a one-way ANOVA in order to detect the effect of the experimental condition (gesture, unrelated, and no gesture) on the four dependent variables (correct recollections, discourse-based inferences, elaborative inferences, and errors). The analysis revealed a main effect of the condition on correct recollections, discourse-based inferences, and errors ($4.07 < F(2, 57) < 9.83$; $.02 < p < .0001$). The effect was not significant for elaborative inferences ($p = .23$). In line with the predictions of the non-deterministic hypothesis, the participants in the unrelated conditions produced correct recollections, discourse-based inferences, and errors to the same extent as participants in the no gesture condition (Tukey post hoc: $.21 < p < 1$). Also, the participants in the gesture condition produced more correct recollections and discourse-based inferences, and fewer errors compared to participants in both the unrelated and no gesture conditions (Tukey post hoc: $.001 < p < .04$).

Discussion of the Results of Experiments 1 and 2

The non-deterministic hypothesis, according to which the uptake of gestural information is not obligatory, predicts that the participants' performance in the unrelated and

in the no gesture conditions should be better in terms of verbatim recognition of sentences (Experiment 1) and worse in terms of correct recollections and discourse based inferences (Experiment 2), compared to performance in the gesture condition. The results of the experiments confirmed such predictions, enforcing the assumption that the uptake of gestural information is not obligatory. Indeed, participants in the unrelated condition, as well as those in the no gesture condition, had better verbatim memory for discourse compared to participants in the gesture condition, at the expense of correct recollections and discourse-based inferences.

Considered as a whole, these results support the non-deterministic hypothesis for a narrative discourse. Within our theoretical framework, the effect of gestures on mental models should be independent of the discourse type. However, some studies in the literature have suggested that narrative contents may have a special status in favoring the construction of a mental model (see, e.g., Zwaan, Langston & Graesser, 1995; Zwaan et al., 1995). To exclude the possibility that the effect of gestures as found in Experiments 1 and 2 may depend upon the fact that the discourse used in those experiments was a narrative discourse with a significant spatial content, we replicated both experiments using a scientific discourse, characterized by a little spatial but high abstract and technical content. Hence, we conducted Experiments 3 and 4 using a discourse on color perception.

Experiment 3: A Replication of Experiment 1 Using a Discourse with a Scientific Content

As for Experiment 1, we expected participants to perform better in recognizing sentences at verbatim level in a discourse accompanied by unrelated gestures (unrelated condition) or delivered without gesturing (no gesture condition), as compared with a discourse

produced with matching gestures (gesture condition). Also, we predict that participants in the unrelated condition and in the no gesture condition have comparable performance.

Participants

Forty-five students from Milan University voluntarily took part in the experiment (28 females and 17 males, mean age: 22). Fifteen were randomly assigned to the gesture condition, fifteen to the unrelated condition and fifteen to the no gesture condition. None of the participants in Experiment 3 had taken part in Experiments 1 or 2.

Materials and Procedures

The experimental material comprised a videotaped discourse presented in Italian by an animated agent. The discourse was part of one used by Cutica and Bucciarelli (2008): it was concerned with the principles of color perception (see Appendix 2A), and lasted about two minutes. The discourse, like the discourse used in Experiments 1 and 2, was presented in three different conditions, for a total of three videotaped fictions. In the gesture condition the animated agent accompanied the discourse with co-speech gestures which were modeled on the basis of co-speech gestures performed by a human actor in the original experiment by Cutica and Bucciarelli (2008).

Two judges examined the gestures produced by the agent and found that he produced 13 representational gestures, 1 deictic gesture, 6 baton gestures and no symbolic gestures. They also excluded the possibility that any of the gestures conveyed information that was implicit or absent in the co-occurring speech. In the unrelated condition, the animated agent's gestures appeared in the same sequence as in the gesture condition, but with a delay of about 20 seconds with respect to the corresponding speech: the gesture track simply moved 20 seconds forward regardless of what was going on in speech. In the gesture and unrelated conditions, the number and type of gestures were the same. Finally, in the no gesture condition the animated agent delivered the discourse without gesticulating. In each condition

the animated agent maintained the same facial expressions and the same labial movements. His voice was the recorded voice of the human actor in the original experiments, and from the beginning and throughout the discourse, the labial movements of the animated agent matched his speech. Figure 2 shows three frames from the videotaped fictions used in the three experimental conditions. The first and second frames depict the gesture produced by the agent while saying the word “water” in the sentence “and blue is refreshing like water” in the gesture and unrelated conditions, respectively. The third frame refers to the animated agent in the no gesture condition while saying the same word. Examples of gestures produced by the actor are in Appendix 2B.

The experimental procedures were identical to those of Experiment 1: participants, randomly assigned to one condition, were invited to watch one of the videotaped fictions, after which they were presented with a list of sentences, one by one in random order. We created 24 sentences, with eight in each category: literally correct, paraphrases, wrong content (examples are included in Appendix 2C).

Results

Several analyses of variance using the Q of Cochran test revealed that the eight stimuli constituting each sentence category were comparable in terms of difficulty ($12.48 < Q < 21.05$; $.88 < p < .43$). Table 3 illustrates the mean correct responses in the three conditions.

(Table 3 about here)

We performed a one-way ANOVA that revealed a main effect of the experimental condition on the ability to correctly identify sentences at verbatim level ($F(2, 42) = 9.07, p = .001$). On the contrary, the effect was not significant for the ability to refute paraphrases ($p = .11$) and wrong content sentences ($p = .28$). In line with the predictions of the non-deterministic hypothesis, participants in both the unrelated (Tukey post hoc: $p < .0001$) and the no gesture conditions ($p = .03$) performed better than participants in the gesture condition

in identifying literally correct sentences. Participants in the unrelated and no gesture condition also showed comparable levels of performance in identifying literally correct sentences, and refuting paraphrases and wrong content sentences (Tukey post hoc: $.11 < p < .66$).

Experiment 4: A Replication of Experiment 2 Using a Discourse with a Scientific Content

In Experiment 4 we expected to replicate the findings that participants in the unrelated condition and in the no gesture condition perform worse than participants in the gesture condition in terms of recollection of correct information and production of discourse-based inferences. We also expected to replicate the finding that participants in the unrelated and the no gesture conditions have comparable performances.

Participants

Forty-five students from Milan University voluntarily took part in the experiment (28 females and 17 males, mean age: 22). Fifteen participants were randomly assigned to the gesture condition, fifteen to the unrelated condition, and fifteen to the no gesture condition. None of the participants in Experiment 4 had taken part in Experiments 1, 2 or 3.

Materials and Procedures

The experimental material was the same as for Experiment 3, namely the videotaped color discourse in the gesture, the unrelated, and the no gesture conditions. The procedures were identical to those of Experiment 2. Again, to ascertain that the actor's gestures were not perceived by the participants as mismatching rather than unrelated, two judges examined each gesture-speech combination produced by the actor in the unrelated condition of the color discourse. They did not classify any of them as gesture-speech mismatches.

To code the results, the discourse was divided into 18 semantic units, corresponding to as many main concepts that the listener could recall (see Appendix 2A). Each concept (i.e. semantic unit) recalled by the participants was evaluated according to the coding schema adopted in Experiment 2.

Results

Two independent judges, blind with respect to the condition they were examining, coded the participants' recollections individually; 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 ($.88 < K < .95$, p always $< .001$). For the final score the judges discussed each item on which they disagreed, until reaching full agreement. Table 4 shows the mean scores for types of recollection in the three experimental conditions.

(Table 4 about here)

As a general result, and in all experimental conditions, scores for correct recollections were higher than for other sorts of recollection. We performed a one-way ANOVA in order to detect the effect of the experimental condition on the four types of recollection. The results revealed a main effect of the experimental condition on correct recollections ($F(2, 42) = 18.01$, $p < .0001$) and discourse-based inferences ($F(2, 42) = 3.68$, $p = .03$). The effect was not significant for elaborative inferences ($p = .81$) and errors ($p = .11$). In line with the predictions of the non-deterministic hypothesis, participants in the unrelated conditions produced correct recollections and discourse-based inferences to the same extent as participants in the no gesture condition (Tukey post hoc: $.37 < p < .91$). Also, participants in both the unrelated and no gesture conditions produced less correct recollections and discourse-based inferences than participants in the gesture condition (Tukey post hoc: $.0001 < p < .04$).

General Discussion and Conclusions

There is a huge body of literature on in-depth investigations into the role of co-speech gestures in discourse comprehension. The focus on gestures whose meaning is either consistent or in contrast with the meaning of the speech they accompany has led to the conclusion that all gestures are however processed. The aim of our study was to verify the determinism of the uptake of unrelated gestural information. In our view, the study fills in a gap by examining whether the uptake of gesture information is mandatory even in cases where gestures are arbitrary with respect to speech. To this aim, the participants in our experiments dealt with an experimental condition (unrelated condition) in which the correspondence in meaning between gestures and speech was disrupted by inserting a 20 seconds delay between the speech and the naturally co-occurring gesture; the delay disrupts the correspondence because the duration of working memory does not exceed 20 seconds. The un-relatedness of the resulting combination gesture-speech was further excluded by two independent judges. Therefore, we are allowed to consider the gestures in the unrelated condition as conveying an arbitrary meaning with respect to the meaning conveyed by speech. As a general result, and against the deterministic hypothesis, gestures unrelated to the speech they accompany produced the same effect as the absence of gestures in both recognition and recall memory tasks. Indeed, the results of Experiments 1 and 3, when taken together, showed that verbatim recall was greater after listening to a discourse accompanied by unrelated gestures or no gestures than a discourse accompanied by gestures. Also, the results of Experiments 2 and 4 showed that a person who listens to a discourse accompanied by unrelated gestures or no gestures at all, will recall less correct information and draw fewer discourse-based inferences compared to a person who listens to a discourse accompanied by

gestures. In particular, the presence of unrelated gestures did not negatively affect participants' performance with respect to the absence of gestures.

Our results enforce the assumption that when gestures are unrelated, the listener does not process their meaning. In particular, they suggest that a mistiming of 20 seconds is sufficient to impede the uptake of gesture information. Indeed, in both recognition tasks, we found that unrelated gestures did not impair verbatim memory for discourse, whereas co-speech gestures did. In both recollection tasks, unrelated gestures did not favor deep comprehension and learning, whereas matching gestures did, although they did not impair comprehension when compared with a situation in which an actor produced a discourse without gesticulating.

Our results are consistent with findings revealing that the uptake of gestural information can be modulated and influenced by some (external) factors. Gullberg and Kita (2009), for example, found that both social (e.g., speakers looking at their own gestures) and physical (e.g., the gesture's location in gesture space) factors, which we kept constant in our experiments, may modulate the uptake of gestural information.

Our results only apparently contrast with those in the literature showing that mismatching gestures are processed together with the speech they accompany (see, e.g., Alibali et al., 1997; Cassell et al., 1999; McNeill et al., 1994; Kelly & Church, 1998; Thompson & Massaro, 1994). Indeed, these studies tested combinations of gesture-speech which however maintain a certain sort of semantic relationship, and we believe that this is why they led to conclude that the uptake of gestures information is mandatory. As the studies in the literature did not consider the case in which gesture and speech are unrelated in meaning, we argued for the necessity to adopt a new paradigm, which albeit exploiting an artificial combination of gesture-speech, disrupts the semantic relationship between gesture and speech, allowing to discover whether gestures' processing is mandatory.

A further case of apparent contrast between our results and those in the literature are the results by Kelly, Ozyurek, and Maris (2010), who used a prime-effect paradigm with different levels of incongruity between gestures and speech and found that the uptake of gestural information is obligatory, and that the strength of the incongruence between speech and gesture affects processing, as strong incongruities impair the integration more than weak incongruities. In their experiment, the higher level of incongruity, i.e., for instance “chop” in speech and “twist” in gesture, is similar to our concept of unrelated gesture. In our view, the difference between their result and ours is due to the experimental material. In particular, the participants in their investigation dealt with simple combinations of gesture and word, whereas the participants in our investigation encountered full discourses. Hence, it is possible that the participants in the unrelated condition of our experiments, at a certain moment from the beginning of the discourse, realized that gestures were incongruent and were therefore able to disregard them.

Our study has three main limits. First, given the nature of the unrelated condition, in which gestures are temporally switched with respect to their natural occurrence, the ecological validity of the results is limited. Second, we did not analyze the co-occurrence of speech with different types of gestures; our experimental design was not apt to investigate deictic, representational, and baton gestures separately. In future, more fine-grained studies might make this kind of investigation possible; for instance, it would be interesting to focus on batons, the gesture category that is least dependent on discourse content, and whose facilitative effect on memory, as pointed out by So, Sim Chen-Hui, and Low Wei-Shan (2012), may rely on partially different cognitive mechanisms. Finally, although we found that gestures are not processed when they are clearly unrelated with speech, our experimental design has not the granularity to shed light on the steps through which the listeners process gesture’s meaning.

Our results suggest to speculate that the participants in our experiments did not monitor the gestures-speech relatedness for the entire length of the discourse; if this would be the case, such monitoring would have resulted in a poorer performance in the unrelated condition compared to the no gesture condition. But future studies could assess more precisely when, in listening to a discourse in which gestures and speech are unrelated, do the listeners begin to neglect gestures' meaning. .

Acknowledgements:

This research was supported by a grant from Italian MIUR to the second author (Prin project 2010-2011, prot. 2010RP5RNM).

References

- 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.
- Bangerter, A. (2004). Using pointing and describing to achieve joint focus of attention in dialogue. *Psychological Science, 15*, 415-419.
- Beattie, G, & Shovelton, H. (1999). Mapping the range of information contained in the iconic hand gestures that accompany spontaneous speech. *Journal of Language and Social Psychology, 18*, 438-462.
- Berger, K.W., & Popelka, G.R. (1971). Extra-facial gestures in relation to speech reading. *Journal of Communication Disorders, 3*, 302-308.
- Bonatti, L. (1998). Possibilities and real possibilities for a theory of reasoning. In Z. Pylyshyn (Ed.), *Essays on representations* (pp. 85-119). Ablex.
- Brown, J. (1958). Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology, 10*, 12-21.
- Bucciarelli, M. (2007). How the construction of mental models improves learning. *Mind & Society, 6*, 67-89.
- Bucciarelli, M., & Cutica, I. (2012). Mental models in improving learning. In N. M. Seel, (Ed.), *Encyclopedia of the sciences of learning, 2012, Part 13*, 2211-2212. Springer: NY.
- Cassell, J., McNeill, D., & McCullough, K.-E. (1999). Speech-gesture mismatches: Evidence for one underlying representation of linguistic and non-linguistic information. *Pragmatics & Cognition, 7*, 1-34.
- Church, R. B., & Goldin-Meadow, S. (1986). The mismatch between gesture and speech as an index of transitional knowledge. *Cognition, 23*, 43-71.

- Church, R. B., Kelly, S. D., & Lynch, K. (2000). Immediate memory for mismatched speech and representational gesture across development. *Journal of Nonverbal Behavior, 24*, 151-174.
- 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. (2012). Mental models in discourse processing. In N. M. Seel, (Ed.), *Encyclopedia of the sciences of learning, 2012, Part 13*, 2213-2215. Springer: NY.
- Ekman, P., & Friesen, W.V. (1972). Hand movements. *Journal of Communication, 22*, 353-374.
- Feyereisen, P. (2006). Further investigation on the mnemonic effect of gestures: Their meaning matters. *European Journal of Cognitive Psychology, 18*, 185-205.
- Galati, A., & Samuel, A. G. (2011). The role of speech-gesture congruency and delay in remembering action events. *Language & Cognitive Processes, 26*, 406-436.
- Garnham, A., & Oakhill, J. (1992). Discourse processing and text representation from a "Mental Models" perspective. *Language & Cognitive Processes, 7*, 204-193.
- 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*, 19-39.
- 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*, 609-640. San Diego, CA: Academic Press.
- Goldin-Meadow, S. (1999). The role of gesture in communication and thinking. *Trends in Cognitive Science, 3*, 419-429.
- Goldin-Meadow, S., & Alibali M. W. (2013), Gesture's role in speaking, learning, and creating language. *Annual Review of Psychology, 64*, 257-283.

- 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., & Sandhofer, C.M. (1999). Gestures convey substantive information about a child's thoughts to ordinary listeners. *Developmental Science, 2*, 67-74
- Graesser, 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.
- Gullberg, M., & Kita, S. (2009). Attention to speech-accompanying gestures: Eye movements and information uptake. *Journal of Nonverbal Behavior, 33*, 251-277.
- Habets, B., Kita, S., Shao, Z., Özyurek, A., & Hagoort, P. (2011). The role of synchrony and ambiguity in speech-gesture integration during comprehension. *Journal of Cognitive Neuroscience, 23*, 1845-1854.
- Hadar, U., Burnstein, A. Krauss, R. M., & Soroker, N. (1998). Ideational gestures and speech: A linguistic investigation. *Language and Cognitive Processes, 13*, 59-76.
- Hadar, U., Wenkert-Olenik, D., Krauss, R. M., & Soroker, N. (1998). Gesture and the processing of speech: Neuropsychological evidence. *Brain & Language, 62*, 107-126.
- Hildebrandt, B., Moratz, R., Rickheit, G., & Sagerer, G. (1999). Cognitive modelling of vision and speech understanding. In G. Rickheit & C. Habel (Eds.), *Mental models in discourse processing and reasoning* (pp. 213-236). Elsevier, North Holland.
- Holle, H., & Gunter, T. C. (2007). The role of iconic gestures in speech disambiguation: ERP evidence. *Journal of Cognitive Neuroscience, 19*, 1175-1192.
- Hostetter, A. B. (2011). When do gestures communicate? A meta-analysis. *Psychological Bulletin, 137*, 297-315

- Iverson, J. M., & Goldin-Meadow, S. (2001). The resilience of gesture in talk: gesture in blind speakers and listeners. *Developmental Science, 4*(4), 416-422.
- 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 & Language, 40*, 577-592.
- 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.
- Kelly, S., Kravitz, C., & Hopkins, M. (2004). Neural correlates of bimodal speech and gesture comprehension. *Brain and Language, 89*, 253-260.
- Kelly, S., Ward, S., Creigh, P., & Bartolotti, J. (2007). An intentional stance modulates the integration of gesture and speech during comprehension. *Brain & Language, 101*, 222-233.
- Kelly, S. D., Ozyurek, A., & Maris, E. (2010). Two sides of the same coin: Speech and gesture mutually interact to enhance comprehension. *Psychological Science, 21*, 260-267.
- Kendon, A. (1983). Gesture and speech: How they interact. In J. A. Wiemann & R. P. Harrison (Eds.), *Nonverbal interaction* (pp. 13-45). Beverly Hills, London: Sage Publications.

- Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press.
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge: Cambridge University 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.
- Krauss, R. M., & Hadar, U. (1999). The role of speech-related arm/hand gestures in word retrieval. In R. Campbell & L. Messing (Eds.), *Gesture, speech, and sign* (pp. 93-116). Oxford: Oxford University Press.
- McNeil, N. M., Alibali M. V., & Evans J. L. (2000). The role of gesture in children's comprehension of spoken language: now they need them, now they don't. *Journal of Nonverbal Behavior*, 24, 131-150.
- McNeill, D. (1985). So you think gestures are nonverbal? *Psychological Review*, 92, 350-371.
- McNeill, D. (1992). *Hand and mind*. Chicago: University of Chicago Press.
- McNeill, D. (2005). *Gesture and thought*. 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.
- Özyürek, A., Willems, R. M., Kita, S., & Hagoort, P. (2007). On-line integration of semantic information from speech and gesture: Insights from event-related brain potentials. *Journal of Cognitive Neuroscience*, 19, 605-616.
- Peterson, L. R., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology*, 58, 193-198.

- Rickheit, G., & Sichelschmidt, L. (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). Elsevier, North Holland.
- Riseborough, M. G. (1981). Physiographic gestures as decoding facilitators: Three experiments exploring a neglected facet of communication. *Journal of Nonverbal Behavior*, 5, 172-183.
- Singer, M. (1994). Discourse inference processes. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 479-515). San Diego: Academic Press.
- Skipper, J., Goldin-Meadow, S., Nusbaum, H., & Small, S. (2007). Speech-associated gestures, Broca's area, and the human mirror system. *Brain & Language*, 101, 260-277.
- So, W. C., Sim Chen-Hui, C., & Low Wei-Shan, J. (2012). Mnemonic effect of iconic gesture and beat gesture in adults and children: Is meaning in gesture important for memory recall?. *Language & Cognitive Processes*, 27, 665-681.
- Thompson, L. A., Driscoll, D., & Markson, L. (1998). Memory for visual-spoken language in children and adults. *Journal of Nonverbal Behavior*, 22, 167-187.
- Thompson, L. A., & Massaro, D. W. (1986). Evaluation and integration of speech and pointing gestures during referential understanding. *Journal of Experimental Child Psychology*, 42, 144-168.
- Thompson, L. A., & Massaro, D. W. (1994). Children's integration of speech and pointing gestures in comprehension. *Journal of Experimental Child Psychology*, 57, 327-354.
- 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. *European Journal of Cognitive Psychology*, 22, 4, 612-639.

- Wang, L., & Chu, M. (2013). The role of beat gesture and pitch accent in semantic processing: An ERP study. *Neuropsychologia*, *51*, 2847-2855.
- Willems, R., & Hagoort, P. (2007). Neural evidence for the interplay between language, gesture, and action: A review. *Brain & Language*, *101*, 278-289.
- Wu, Y. C., & Coulson, S. (2005). Meaningful gestures: Electrophysiological indices of iconic gesture comprehension, *Psychophysiology*, *42*, 654-667.
- Wu, Y. C., & Coulson, S. (2010). Gestures modulate speech processing early in utterances. *Cognitive Neuroscience & Neuropsychology*, *21*, 522-526.
- Zwaan, R. A., Langston, M. C., & Graesser, A. C. (1995). The construction of situation models in narrative comprehension: An event-indexing model. *Psychological Science*, *5*, 292-297.
- Zwaan, R. A., Magliano, J. P., & Graesser, A. C. (1995). Dimensions of situation model construction in narrative comprehension. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *21*, 386-397.
- Zwaan, R. A., & Radvansky G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, *123*, 162-185.

Appendix 1

Appendix 1A The funfair discourse used in Experiments 1 and 2 (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"./

Appendix 1B Examples of co-speech gestures produced by the animated agent in both the gesture condition and the unrelated condition (Experiments 1 and 2 - Funfair discourse).

The left and the central columns present the co-occurrence of speech and gestures in the gesture condition, whereas the right and the central columns present the co-occurrence of speech and gestures in the unrelated conditions. Please note that the beginning of each condition is highlighted in bold.

Underscoring of semantic units indicates the duration of the gesture with the corresponding number.

Semantic units Gesture condition	Computer animated agent's gestures	Semantic units Unrelated condition
BEGINNING OF GESTURE CONDITION FICTION		
It was there, at the funfair, it was there that I found her,	Hands still, resting on thighs.	
and it was at the funfair that I lost her.	Hands still, resting on thighs.	
<u>It was a vast funfair.</u> ¹	1) Raises both hands simultaneously to a position at the level of the navel, palms down. From here, the hands are raised to head height in a continuous movement, curving outwards as if describing a circle, with palms facing forward towards the hypothetical interlocutor. Both hands return to chest level.	<u>The candy floss scented the air,</u> <u>and she was eating it, and she blew on her candy</u> ¹
<u>A funfair with shooting-ranges</u> ² <u>and candy floss stalls,</u> ³	2) The right hand closes into a fist, and the left hand remains open with palm facing to the right; then both hands open and assume a position as if to point to the agent's left; the left hand is slightly higher than the right; the left arm is almost fully straightened, while the right one is more bent. Then both hands move down to the thighs. 3) The right hand is rapidly raised to stomach level, outside the line of the body; the hand is half open as if to indicate a direction; it is then quickly moved back to rest on the thigh.	<u>And I was all covered with white powder</u> ² <u>She started laughing</u> ³

<p><u>and Japanese bagatelle tables⁴ stalls with bottles of champagne,⁵</u></p>	<p>4) The right hand is raised from the right leg and moves to one side of the body (further from the body than in the previous movement), almost to stomach height. The hand is half open, with the fingers towards a hypothetical interlocutor, as if to indicate a direction; the arm is partly bent.</p> <p>5) Just before the right hand comes to rest on the leg, the left hand is raised from the other leg and moves forward; the thumb and index finger are extended and pointing, while the other fingers are partly bent and the palm is almost hidden by the fingers. The hand then returns directly to the left leg.</p>	<p><u>and I asked her: "What's your name?"⁴</u></p> <p><u>And she shouted to me, "I'll⁵ tell you later"</u></p>
<p><u>and showmen's booths⁶ and roundabouts.⁷</u></p>	<p>6) Both hands are raised simultaneously from the legs. The right hand is open with the palm towards the right leg, away from the body, with the thumb pointed upwards; the left hand is open and moves away from the body, with the palm still facing the thigh, and the index finger and thumb straight. Both hands then move back to the thighs.</p> <p>7) The left hand is raised immediately from the left leg and moved forward and to the right, showing the palm with fingers open, to chest height (the elbow is bent at 90°), and remains in that position for 2 seconds; the right hand is open and remains still at hip height, with the palm towards the thigh.</p>	<p>BEGINNING OF UNRELATED CONDITION FICTION</p> <p><i>(The actor is silent)⁶</i></p> <p><u>It was there, at the funfair⁷</u></p>
<p><u>And the roundabouts turned and creaked,⁸</u></p>	<p>8) Both hands start to move from left to right, held forwards at chest height; the right hand continues the movement gradually downwards to hip level. Then both hands are moved back towards the legs but do not come to rest on them. The left hand, open and palm facing downwards, slides downwards and outwards from the left leg, with the arm extended. In a continuous movement, the hand moves back upwards in a circular fashion and comes back to rest on the left thigh.</p>	<p><u>it was there⁸ that I found her,</u></p>
<p><u>and the candy floss scented the air,⁹</u></p>	<p>9) Both hands are moved away from the legs up to the mouth area, in a cupped position 2-3 cm apart with the palms towards the mouth and fingers slightly bent, moving rapidly backwards and forwards for around 2 seconds.</p>	<p><u>and it was at the funfair that I lost her⁹.</u></p>
<p><u>and the rifles shot.¹⁰</u></p>	<p>10) Then the open right hand moves to the chest area, while the left hand moves down to hip level. The left hand, open and palm upwards, is raised in a circular movement, and then from the hip area it is moved up to the chest, near to the right hand.</p>	<p><u>It was a vast funfair¹⁰</u></p>
<p><u>I was shooting at the target.¹¹</u></p>	<p>11) Both hands are at chest height, with palms facing each other; both hands are moved towards the left. The left hand moves away from the body, while the right remains at stomach level. The ring and little fingers of the left hand close, while the other digits are straight; at the same time, the middle, ring and little fingers of the right hand close, while the thumb and index finger remain straight and extended.</p>	<p><u>A funfair¹¹ with shooting-ranges¹² and candy floss stalls.¹³</u></p>

<p>I can <u>shoot at the target very well</u>¹² and I am <u>proud of it</u>.¹³</p>	<p>12) From that position, the hands are raised momentarily to shoulder height, with index fingers pointing upwards, and then immediately lowered to their initial position. The right hand is open, with the palm towards the stomach and the arm bent; the left hand is at the front, with thumb and index finger straight, while the other fingers and the right arm are slightly bent. 13) The hands are brought up to cover the face for 2 seconds; the hands are open, with the palms towards the face.</p>	
<p>No, no,¹⁴ wait a moment¹⁵</p>	<p>14) The hands remain in this position for 2 seconds and then slowly descend to hip level. When the hands reach stomach level, the right hand moves towards the left, while the left hand opens and moves forwards, with the fingers almost fully extended. Then the hands move directly towards the thighs, but do not actually come to rest on them. 15) The left hand is raised quickly up to the top of the chest, with the palm towards the right and fingers open. The right hand is raised slightly, up to hip level, with the palm still facing downwards. The fingers are semi-extended. Both hands move rhythmically at around hip height and are brought slightly outwards.</p>	<p>and Japanese <u>bagatelle tables</u>¹⁴ <u>stalls with bottles of champagne</u>.¹⁵</p>
<p>I am wrong!</p>	<p>The hands stay still near the legs for 7 seconds.</p>	<p>and showmen's booths and roundabouts.</p>

Appendix 1C Examples of sentences used for the Recognition Task of Experiment 1

Literal I can shoot at the target very well

Paraphrases I am really able to shoot at the target

Wrong I do not know how to shoot at the target

Literal I shot, the egg popped up. I turned aside, she wasn't there any more.

Paraphrases When I shot at the egg she disappeared.

Wrong While I was going to shoot to the egg, I turned aside and she wasn't there any more

Appendix 2

Appendix 2A The color discourse used in Experiments 3 and 4 (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./ The theory of association, however, is not more interesting or prolific in this field than in others./ In addition, the effects of colors are too direct and spontaneous/ to be simply the result of an interpretation given through knowledge./ On the other hand, no hypothesis has been advanced so far on the kind of physiological process/ that could help to explain the influence of colors on the organism./ The need to discuss the form makes us feel on more solid ground, though,/ as we can compare the expression of specific patterns/with that or more general properties/ such as spatial orientation, balance or the geometrical characteristics of the outlines.

Appendix 2B Examples of co-speech gestures produced by the animated agent in both the gesture condition and the unrelated condition (Experiments 3 and 4 – Color discourse)

The left and the central columns present the co-occurrence of speech and gestures in the gesture condition, whereas the right and the central columns present the co-occurrence of speech and gestures in the unrelated conditions. Please note that the beginning of each condition is highlighted in bold. Underscoring of semantic units indicates the duration of the gesture with the corresponding number.

NB: if the numbering of the gestures corresponding to the semantic units is not perfectly sequential, this is likely to be due to changes in the sentence structure in the translation from the Italian. We have therefore preferred to maintain the speech-gesture correspondence rather than the sequential gesture numbering.

<p>Semantic units Gesture condition</p>	<p>Computer animated agent’s gestures</p>	<p>Semantic units Unrelated condition</p>
<p>BEGINNING OF GESTURE CONDITION FICTION</p> <p>It’s beyond dispute <u>that colors carry strong expressive components</u>¹</p> <p><u>Some attempts have been made to describe the specific expressive characteristics of the various colors</u>²</p>	<p>1) The agent’s hands are resting on his legs at knee height. He raises both hands simultaneously to the level of the navel. From here, the hands open outwards with palms facing slightly forwards, towards the hypothetical interlocutor. The hands are then brought swiftly back to the thighs.</p> <p>2) The left hand is raised from the left leg up to the mouth area, half open and with the palm towards the right. The hand moves down to chest level in a continuous movement and then assumes an up-down undulating motion. Initially, the fingers are all partly closed, then the index finger gradually extends almost completely; the palm of the hand faces towards the interlocutor (first at mouth level and then at chest height). At the same time, the open right hand is raised from the right leg to a position at the level of the navel, with the palm facing</p>	<p>There is a very widespread belief that <u>the expression of colors is based on association</u>¹.</p> <p>Therefore, <u>red should be considered exciting</u>,²</p>

GESTURAL INFORMATION UPTAKE

	to the left and fingers half extended.	
<u>and to draw some general conclusions from the symbolic use the different cultures have made of them.</u> ³	3) From the chest, both hands move down onto the legs and closer to each other (to about 2-3 cm apart), with palms facing each other, fingers extended and thumbs pointing upwards. The right hand is raised from the right leg up to chest level, with fingers open, while the left hand rises at the same time in line with the right hand, first with fingers half closed and then extended. The right hand is then brought back onto the right leg, with palm facing to the left and thumb raised, while the left hand moves in small circles, with fingers semi-extended. Finally, the left hand moves back onto the left leg, with the palm facing to the right and fingers extended.	<u>because it reminds us of the connotations of fire, blood and revolution</u> ³ .
There is a very widespread ⁶ belief ⁴ that the expression of colors is based on association ⁵ .	4) The left hand is raised from the left leg up to neck height, with the index finger pointing upwards and the other fingers half closed, whereas the right hand remains open at hip level, with the palm facing to the left. 5) The left hand moves down from the mouth to shoulder height, then in an up-down waving motion, first with the hand open and then with fingers half closed, with the palm facing downwards and to the right. 6) The right hand is raised from the right thigh up to the same height as the left hand; both palms are parallel to the thighs. The hands move forwards, and both palms turn towards the agent; the arms remain partly bent.	<u>Green</u> ⁴ evokes the <u>restorative thought of nature</u> ⁵ ,
		<u>and blue</u> ⁶
<u>Therefore, red</u> ⁷ <u>should be considered exciting.</u> ⁸	7) Both hands are brought up to chest height, first closer together then further apart again. 8) With both hands open at chest level, the right hand moves closer to then further from the left hand.	<u>is refreshing like water</u> ⁷ .
<u>because it reminds us of the connotations of fire</u> ⁸ , <u>blood</u> ⁹ <u>and revolution</u> ¹⁰ .	8) From the chest, both hands are moved to the shoulder area, with palms facing each other and fingers extended. They are then quickly lowered to waist level, with fingers closed. The fingers immediately reopen and the right hand is raised up to the right shoulder; the fingers open and the index finger points upwards above the shoulder, with the palm turned slight to the left. The left hand stops at chest height, facing forwards with the palm open. 9) The left hand moves up and down repeatedly, while the right hand remains at shoulder level. 10) Both hands are moved to chest level; the fingers are open and the hands extended forwards with palms facing the interlocutor. The index fingers appear to point upwards.	<u>The theory of association, however,</u> ⁸ <u>is not more interesting or prolific</u> ⁹ <u>in this field</u> ¹⁰ <u>than in others.</u>
<u>Green</u> ¹¹ <u>evokes the restorative thought of nature</u> ¹² ,	11) The left hand is lowered slightly and turned, in an open position, so that the palm faces upward, with the thumb and index finger more extended than the other digits. It then turns again, so as to finish with the palm facing the right hand. 12) Both hands are turned towards the agent and slightly folded (i.e. with the fingers of the right hand covering those of the left). Both hands then move outwards, with a turn of the arms, until the right hand is outside the line of the body, while the left hand remains slightly closer to the body. The hands return to chest level.	<u>In addition,</u> ¹¹ <u>the effects of colors are too direct and</u> ¹² <u>spontaneous</u>

GESTURAL INFORMATION UPTAKE

<p><u>and blue¹³ is refreshing like water¹⁴.</u></p>	<p>13) The right hand moves down from the chest onto the right thigh, while the left hand is raised to shoulder level. 14) The left hand is moved repeatedly up and down, moving the fingers at the same time.</p>	<p><u>to be simply the result¹³ of an interpretation given¹⁴ through knowledge.</u></p>
<p><u>The theory of association, however, is not more interesting¹⁵ or prolific¹⁶ in this field than in others.</u></p>	<p>15) The left hand moves down from the shoulder to the chest and the palm is turned towards the interlocutor, with fingers half closed. Simultaneously, the right hand is raised from the right leg up to the same height as the left hand, with the palm facing towards it. 16) Both hands close and are then moved from the chest to rest on the legs.</p>	<p>BEGINNING OF UNRELATED CONDITION FICTION <u>It's beyond dispute¹⁵</u> <u>that colors carry strong expressive components¹⁶</u></p>
<p><u>In addition¹⁷, the effects of colors are too direct and spontaneous¹⁸</u></p>	<p>17) The hands are raised to hip level, with palms facing each other and fingers open. 18) The right hand remains at hip level, open and with the palm facing left, and moves outwards away from the body. The left hand is raised to chest height, outside the line of the body, and is pushed quickly forwards, with fingers open, and then pulled back equally quickly towards the chest 19) The right hand is raised from the right leg by about 5 cm, with the palm towards the leg, while the left hand remains at chest level</p>	<p><u>Some attempts have been made¹⁷ to describe</u> <u>the specific¹⁸ expressive characteristics of the various colors¹⁹</u></p>
<p><u>to be simply the result¹⁹ of an interpretation given through knowledge²⁰</u></p>	<p>20) The right hand moves to around stomach level, away from the body; the hand is open with the index finger almost fully extended, the other fingers half closed, and the palm facing left. It is then brought back to rest on the leg, closing almost into a fist. At the same time, the left hand also moves slightly towards the side of the body, then returns to its previous position and then moves towards the side again</p>	<p><u>and to draw some general conclusions from the symbolic use the different cultures have made²⁰ of them</u></p>

Appendix 2C Examples of sentences used for the Recognition Task of Experiment 3.

Literal Green evokes the restorative thought of nature.

Paraphrases The color evoking the restorative thought of nature is green.

Wrong Green evokes the sense of tiredness of nature.

Literal Blue is refreshing like water.

Paraphrases Blue color gives a feeling of freshness like water.

Wrong Blue color is fresh like sea.

Table 1. Mean correct performance with the different sorts of sentences in the three experimental conditions in Experiment 1.

	Literally correct (<i>n</i> = 8)	Paraphrases (<i>n</i> = 8)	Wrong content (<i>n</i> = 8)
Condition			
Gesture (<i>n</i> = 20)			
<i>M</i>	4.65	6.90	4.70
<i>SD</i>	1.35	1.20	2.06
Unrelated (<i>n</i> = 20)			
<i>M</i>	6.15	6.00	5.65
<i>SD</i>	.99	1.45	1.35
No Gesture (<i>n</i> = 20)			
<i>M</i>	5.60	6.15	5.55
<i>SD</i>	.94	1.18	1.19

Table 2. Mean types of recollections in the three conditions of Experiment 2.

Condition	Correct recollections	Discourse-based inferences	Elaborative inferences	Errors
Gesture ($n = 20$)				
<i>M</i>	9.45	.40	.10	.10
<i>SD</i>	1.82	.50	.31	.31
Unrelated ($n = 20$)				
<i>M</i>	7.10	.10	.15	.50
<i>SD</i>	2.05	.31	.37	.61
No Gesture ($n = 20$)				
<i>M</i>	7.30	.10	0	.50
<i>SD</i>	1.69	.31	-	.51

Table 3. Mean correct performance with the different sorts of sentences in the three experimental conditions in Experiment 3.

	Literally correct (<i>n</i> = 8)	Paraphrases (<i>n</i> = 8)	Wrong content (<i>n</i> = 8)
Condition			
Gesture (<i>n</i> = 15)			
<i>M</i>	3.40	5.53	4.27
<i>SD</i>	1.40	1.25	1.71
Unrelated (<i>n</i> = 15)			
<i>M</i>	5.27	4.93	4.67
<i>SD</i>	0.96	1.03	1.35
No Gesture (<i>n</i> = 15)			
<i>M</i>	4.60	5.73	5.13
<i>SD</i>	1.24	0.88	1.30

Table 4. Mean types of recollections in the three conditions of Experiment 4.

Condition	Correct recollections	Discourse-based inferences	Elaborative inferences	Errors
Gesture ($n = 15$)				
<i>M</i>	5.53	.53	.13	.20
<i>SD</i>	1.36	.52	.35	.41
Unrelated ($n = 15$)				
<i>M</i>	3.00	.20	.13	.53
<i>SD</i>	.93	.41	.35	.52
No Gesture ($n = 15$)				
<i>M</i>	3.60	.13	.07	.53
<i>SD</i>	1.30	.35	.29	.52

Fig. 1. Frames from the *Funfair discourse* video of Experiments 1 and 2. They depict the agent's gestures while saying the word "big" of the sentence "It was a very big funfair". Frame (a) belongs to the gesture condition, frame (b) to the unrelated condition, and frame (c) to the no gesture condition.



Frame (a)



Frame (b)



Frame (c)

Fig. 2. Frames from the *Color discourse* video of Experiments 3 and 4. They depict the agent's gestures while saying the word "water" of the sentence "and blue is refreshing like water". Frame (a) belongs to the gesture condition, frame (b) to the unrelated condition, and frame (c) to the no gesture condition.



Frame (a)



Frame (b)



Frame (c)