How iconic gestures and speech interact in the representation of meaning: Are both aspects really integral to the process?

JUDITH HOLLER and GEOFFREY BEATTIE

The spontaneous movements of the hands that we produce while we are speaking have become a prime focus of research over recent decades. A number of researchers have proposed schemes as to how to classify these hand movements into different types of gestures according to the different functions that they may fulfill in interpersonal communication (see, for example, Efron 1941, Ekman and Friesen 1969, McNeill 1992). In this article, we are concerned with investigating the communicational role of one particular type of those hand gestures, namely iconic gestures. The term *iconic gestures* (McNeill 1985) refers to hand gestures that represent meaning that is closely related to the semantic content of the segments of speech that they accompany. A frequently cited example is the following:

'and she [chases him out again]'
[hand appears to swing an object through the air] (McNeill 1992: 13–14)

In this example, the extract of speech and the iconic gesture represent information relating to the same action. Thus, McNeill concluded that gesture and speech are semantically parallel in that they refer to the same underlying cognitive representation. At the same time, however, he argued that 'along with this kind of co-expressiveness, there is also complementarity' (McNeill 1992: 13). This is because both communicational channels reveal somewhat different aspects of the same scene so that the representations of semantic information in both communicational channels complement each other to convey the full meaning of a single cognitive representation. Thus, in the example above, 'the gesture reveals the weapon while speech conveys the action (chasing) and the idea of recurrence' and 'jointly, speech and gesture give a more complete insight into the speaker's thinking' (McNeill 1992: 13) in that both reveal different semantic aspects of the same scene. In short, McNeill's theory seems to predict that the gestural and the verbal representation overlap only concerning their semantic reference point, but that they are often complementary concerning the actual representation of the various semantic aspects.

Without the simultaneous presence of a gesture, a conversational partner would lack a considerable amount of information, and it follows therefore, according to McNeill, that iconic gestures are *integral* to the process of the communication of meaning.

McNeill's theory of iconic gestures was seminal in effecting a new conceptualization of interpersonal communication. It argued that we have to move away from the theoretical position that focuses on speech as the channel that represents semantic information, and nonverbal behavior, including gestures, as the channel that conveys information about all the other important aspects of communication, such as interpersonal attitudes and emotion (see, for example, Tower, Bryant, and Argyle 1978: 42, but also Ellis and Beattie 1986: 165). According to McNeill, iconic gestures are not part of nonverbal communication at all; rather they communicate semantic information about the world like speech itself.

However, we need to evaluate critically a number of core points of McNeill's theory at this stage. McNeill described the interaction of the gestural and the linguistic systems by saying that 'speech and gesture must cooperate to express the person's meaning' (McNeill 1992: 11). The problem is here that the term *cooperation* is somewhat flexible, because by definition, cooperation means *working together to the same end.* Such a definition does not determine the size of the contributions of the individual components that cooperate together actually. It may be, for example, that their contributions are equal, or that component A, the speech, contributes a large part, whereas component B, the gesture, contributes a much smaller part, or it could be the converse. Furthermore, with regard to McNeill's description of how iconic gestures and speech interact in the representation of semantic information, we can find further evidence of vagueness. What exactly does McNeill mean when he says that gesture exhibits 'many details' that are not exhibited in speech or 'images that cannot always be expressed in speech' (see McNeill 1992: 11)? Which semantic details and what kinds of images is he actually referring to? Is there any general pattern in the representation of semantic information that characterizes the interaction of gesture and speech generally? Thus, we would argue that although McNeill's communication theory of iconic gestures represents an important theoretical advance in our understanding of gestures, it is, however, characterized by a lack of detail in its overall specification.

Experimental investigations advancing McNeill's theory of iconic gestures

Beattie and Shovelton (1999a, 1999b, 2002) attempted to develop McNeill's theory and empirically investigated the communicational role of

iconic gestures by applying what they called a semantic feature approach, despite the fact that some of the semantic categories they applied were quite low level categories like 'movement', 'size', or 'direction', and some were much higher level, more complex categories like 'identity' and 'description of action'. All of their studies were decoder based and involved questioning their participants about the kinds of semantic information that they were able to glean from video clips, which showed individual iconic gestures (either with or without the accompanying speech), or from the clausal units of speech that were originally accompanied by the iconic gestures. Beattie and Shovelton found that iconic gestures communicate considerable amounts of semantic information when they occur together with speech and that iconic gestures also communicate considerable amounts of semantic information in the absence of speech. In particular, their decoders gleaned a significant amount of information about 'relative position' (the position of moving or stationary objects relative to anything else, see Beattie and Shovelton 1999b) in all of their studies (1999a, 1999b, 2002) and about 'size', of any objects identified, in the majority of their studies (1999a, 1999b, 2002 study a). Although Beattie and Shovelton's work can be seen as developing the communication theory of iconic gestures considerably in that they determined exactly which kinds of semantic information iconic gestures communicate, their work has also a number of specific limitations (see Holler and Beattie 2002).

Holler and Beattie (2002) tried to overcome some of these shortcomings, first, by investigating the interaction of the gestural and the linguistic systems using gestural and verbal material that had been generated by participants engaging in conversational interaction with another person, whereas Beattie and Shovelton's participants narrated cartoon stories essentially without interacting with an interlocutor. Research by Beattie and Aboudan (1994), however, suggests that the frequency of gestures increases considerably in interaction compared to monologue conditions. This means that a relatively high proportion of participants were excluded from Beattie and Shovelton's studies because they did not generate any iconic gestures (in their paper from 1999a, Beattie and Shovelton stated that only 50 percent of their participants used iconic gestures, whereas a mere 13.2 percent of the participants did not generate any gestures in Holler and Beattie's study [2002]). This, of course, could affect the ecological validity of their data, because the iconic gestures that represented the basis for Beattie and Shovelton's first study derived from only seven different narrators. Furthermore, because the social context seems to have a crucial impact on the production of gestures, one might also argue that this may affect not only the frequency of gestures, but also potentially other aspects like the quality of the gestures. Beattie and Shovelton investigated

the kinds of semantic information communicated by iconic gestures generated during monologue, rather than the kinds communicated by iconic gestures generated in the more natural context of dialogue.

Second, Holler and Beattie used ideational units (see Butterworth 1975) instead of clausal units of language. McNeill based his theoretical arguments on syntactic clauses, and such units also formed the basis of Beattie and Shovelton's research. Although the iconic gestures chosen for their analyses might sometimes have referred to the same semantic idea as a whole sentence or a group of sentences, Beattie and Shovelton (1999b, 2002) analyzed only the clausal unit that was directly accompanied by the iconic gesture. The reason for this was McNeill's argument that iconic gestures almost never cross clause boundaries. However, an analysis using strictly clausal units may exclude important verbal information relevant to the semantic idea, but represented outside the clausal unit selected for the analysis. This is because an essential characteristic of verbal communication is that it unfolds over time, and thus semantic ideas can unfold over stretches of speech larger than individual clauses.

Thirdly, the semantic categories that Beattie and Shovelton applied were themselves rather broad. For example, the semantic category 'relative position', which appears throughout the Beattie and Shovelton studies, can be deconstructed into sub-categories, such as 'relative position: agent-object', 'relative position: agent-instrument', 'relative position: object-instrument', and 'relative position: object-surrounding space', and the semantic category 'size' can, for example, be split up into three sub-categories, namely 'size: agent', 'size: object', and 'size: instrument' (see Holler and Beattie 2002). Deconstructing the categories in this manner, Holler and Beattie found that the conclusions look somewhat different from those of Beattie and Shovelton. Information relating to the first two position-subcategories was found to be almost exclusively represented by iconic gestures, whereas the latter two were almost exclusively represented by speech. Information about size was found to be exclusively represented by speech apart from one exception. Thus, we would argue that it is essential to investigate communicational aspects of iconic gesture and speech from a more micro-analytic perspective.

The present study

The study by Holler and Beattie (2002), however, was also somewhat limited in scope. Deriving, as it did, from Beattie and Shovelton, Holler and Beattie included only those semantic categories that had already been shown to be significantly communicated by iconic gesture, namely 'relative

Table 1. *The semantic categories applied in the present study*

Entity	1. Agent
	2. Object
	3. Instrument
Action	4. Body-parts involved in the movement
	5. Direction of the movement
	6. Point of contact defined
	7. Force
Relative position	8. Relative position: Agent – Object
	9. Relative position: Agent – Instrument
	10. Relative position: Object – Instrument
	11. Relative position: Object – Space
Size	12. Size: agent
	13. Size: object
	14. Size: instrument
Shape	15. Agent
	16. Object
	17. Instrument
Shape of a part	18. Agent
	19. Object
	20. Instrument

position' and 'size'. However, to be able to draw broader conclusions about how iconic gesture and speech interact in the communication of meaning, we must extend the original analysis to a fuller range of semantic categories. Starting with the present corpus, we empirically derived 20 semantic categories, including the semantic categories 'relative position' and 'size' with their seven subsidiary categories as applied in Holler and Beattie's previous study (see Table 1; for a more detailed definition of the individual semantic categories, please see the 'Method' section). These 20 semantic categories were considered to offer themselves a fuller descriptive base to capture the range of semantic information contained in the iconic gestures and in the speech extracts from the present corpus.

We should emphasize that McNeill (1985, 1992) did not apply any such semantic categories in his discussion of gesture and speech interaction.

Apart from the fact that the present study includes the full range of semantic categories instead of only the categories 'position' and 'size' considered in Holler and Beattie (2002), it also differs from the earlier study in that it uses a revised version of the mathematical scheme for coding the gestural and verbal material. Originally, the coding scheme consisted of three informational values, 0, 0.5, and 1 for both gesture and speech. The

informational value 0 was given when a gesture or an extract of speech contained no information at all concerning a certain semantic feature, the informational value 1 was given when the information was clear information, and the informational value 0.5 was given when information was considered as 'unclear' or 'ambiguous'. The procedure for scoring involved two steps: First, it was decided, from an analyst's point of view, whether certain semantic information was represented by a gesture or an extract of speech (if not, the score was 0). Second, if the information was represented, it was decided whether the information could be perceived as unambiguous or ambiguous from an interlocutor's point of view in the absence of the respective other communicational channel (respectively, an informational value of 1 or 0.5 was given). However, the difficulty of this scoring scheme was that it combined two quite different perspectives, that of an analyst (the score of a provisional 0 or 1) and that of a decoder (with 0.5 meaning unclear and ambiguous, 1 meaning clear and unambiguous).

However, here we take a somewhat different line. In order to investigate the kind of semantic information that is encoded in gesture and speech and how both channels interact (as opposed to what kind of semantic information iconic gestures communicate over and above speech), it may be much more sensible to take into account that gesture and speech are two communicational channels that operate under very different conditions. Speech is principally designed to communicate information in the absence of iconic gestures (for example, many people do not gesture at all or only very rarely, and concerning those people who do gesture, only a relatively small proportion of their speech is accompanied by iconic gestures, and finally, speech communicates information in the absence of gesture, for example when we talk on the phone). Thus, speech, to a large extent, can be interpreted reliably by someone who does not receive any additional information from other communicational channels such as gesture. Iconic gestures, on the other hand, are reliably interpretable only with regard to the accompanying speech, with, according to Beattie and Shovelton (1999b, 2002), information about relative position and information about unmarked size representing perhaps the exceptions. Consequently, the revised criteria for scoring take this crucial difference into account. This means that iconic gestures are interpreted with regard to the accompanying speech as well as against the background of the source of narration (the reason for this is that, in line with the hypothesis of the present study, a gesture might represent certain semantic information about the original scene that is not represented by speech). With regard to the data that is being analyzed in the present study, this resulted in a binary coding scheme for iconic gesture consisting only of the informational values 0 (which

indicates that a certain semantic feature is not represented by the iconic gesture) and 1 (which indicates that a certain semantic feature is represented by the iconic gesture), because all the gestural representations happened to be relatively clearly interpretable; thus, an informational value of 0.5 that could potentially be considered for scoring 'ambiguous gestures' was not necessary.

The scoring criteria for speech changed in the concern that the verbal information is analyzed solely by taking into account no information other than that contained in the linguistic unit, apart from the fact that the context in which the utterance occurred was a 'Tom & Jerry' cartoon story. In their previous study, Holler and Beattie were less consistent concerning context information. With regard to information about the size of the entities, the immediate context was taken into consideration. The pronoun 'him', for example, was considered as providing clear information about the identity (and thus about the size) of a character, because pronouns are only used when a narrator assumes that his or her conversational partner knows from the directly preceding context to whom or what the narrator is referring. With regard to any other information, however, the immediate context has not been taken into account, but was considered from the perspective of a completely naïve interlocutor. To eradicate this methodological shortcoming, the present analysis considers no context information apart from the fact that the context of the narrative is a 'Tom & Jerry' cartoon (and thus any interlocutor's knowledge about a cartoon world). The reason for this is that, for the experimental situation, the narrators were told that they should imagine the experimenter as a naïve interlocutor who does not know anything about the source of narration except that it is a 'Tom & Jerry' cartoon. The fact that the narrative is about a cartoon story, and which type of cartoon story, were thus the only parameters that set the semantic frame for the conversational interaction between narrator and interlocutor, and thus the determinants of the conditions under which the verbal and gestural information has been encoded.

Furthermore, the criteria for scoring verbal information have been revised and slightly modified in that they were defined more precisely with regard to the inferences that interlocutors make due to the fact that 'people, in speaking and writing, consistently leave out information that they feel can easily be inferred by the listener or reader' (Schank and Abelson 1977: 22), and which can thus be considered as an important characteristic of human communication. For example, with regard to the phrase 'he's holding a spike to stab Tom', an interlocutor will infer readily that the agent is holding the spike in the hand(s), although the hands are not explicitly mentioned. With regard to how the agent and the instrument

are positioned relative to each other, however, it can be assumed that an interlocutor makes no inferences, or only weak inferences, because there are considerably more possibilities for inferences (left, right, behind, in front, etc.). Thus, one additional aspect that we took into account here is the assumed strength of certain inferences from the linguistic channel.

Consequently, the scoring scheme for speech remains as a tripartite scheme, and using revised criteria for scoring, it differs between explicitly represented semantic information (informational value = 1), implicitly represented information, assuming relatively strong inferences (informational value = 0.5), and no information or relatively weak inferences assumed (informational value = 0). For a more detailed definition of the scoring criteria and the individual informational values, please see the 'Method' section.

In conclusion, the present investigation of the interaction of the gestural and the linguistic systems goes beyond previous research in a number of ways. First, it goes beyond McNeill's analysis in that it is based on an extensive data set, whereas McNeill's arguments seem to derive from individual examples. Second, the gestural and verbal material analyzed in the present study has been generated during conversational interaction rather than during monologue. Third, to obtain findings of increased ecological validity, the present study is based on the analysis of ideational units, which may be considerably larger than individual clauses. Fourth, we apply here a semantic feature approach, which consists of a wide range of semantic categories that are suitable to capture the information that has been generated in a variety of narratives, and which provide a substantial basis for investigating how gesture and speech interact in the representation of information relating to these core semantic features.

Method

Participants

Thirty-eight students of the University of Manchester took part in the experiment. All the participants were native English speakers.

Stimuli

The stimulus material consisted of two 'Tom & Jerry' cartoon stories (Kelleher 2000), which covered two A4 sheets each.

Procedure

The experiment involved participants narrating cartoon stories while interacting with one of the experimenters (JH), who asked them questions about the actions and the characters in the cartoons. The experimenter sat on a chair, facing the participant but with her back against the wall onto which the cartoon stories were projected. The participants were told that they should imagine the experimenter being a person who had never seen these pictures of the cartoon and who did not know the contents of the stories apart from the fact that the narrative is about a 'Tom & Jerry' cartoon story. Thus, the situation was supposed to resemble everyday conversations, which generally involve face-to-face interaction with one conversational partner telling another person something that he or she did not know, but which require a certain basis of mutual knowledge or basic context information that they share (see, for example, Clark 1996).

   The conversations took place as individual trials, and each participant was filmed by an unobtrusive camera. The participants were informed about the presence of the camera before the experiment (see British Psychological Society Guidelines 2000), but they were told that the study investigated how well people could narrate cartoon stories, in order to prevent the participants from focusing on nonverbal communication. After the experiment, the participants were fully debriefed and thanked. Furthermore, before the start of the experiment, each participant was informed about his/her right to withdraw from the experiment at any time before signing the final consent from, which was handed to him/her after he/she had been debriefed.

Definitions

The following are definitions of the semantic categories applied to the present study:

1. *Entity: agent* — This category refers to the information that an agent is an integral part of the semantic event.
2. *Entity: object* — This category refers to the information that an object, that is the recipient of an action, is an integral part of the semantic event.
3. *Entity: instrument* — This category refers to the information that an instrument, in terms of a third entity, is an integral part of the semantic event.
4. *Action: body-parts* — This category refers to information about the body-parts that the agent uses to accomplish an action.

5. *Action: direction* — This category refers to information about the direction of the movement that is critical with regard to the semantic event.
6. *Action: point of contact* — This category refers to information about the exact point on the object (recipient of the action) that is contacted by the movement.
7. *Action: force* — This category refers to the information as to whether a particular amount of force or a particular lack of force is associated with the action.
8. *Relative position: agent – object* — This category refers to information about how agent and object are positioned relative to each other in terms of the object being in front of the agent, behind the agent, at the agent's left side, at the agent's right side, below the agent, or above the agent.
9. *Relative position: agent – instrument* — This category refers to information about the relative position of agent and instrument in terms of the instrument being in front of the agent, below the agent, or above the agent, to the agent's left, or to the agent's right.
10. *Relative position: object – instrument* — This category refers to information about the relative position of object and instrument in terms of the instrument being in front of the object, behind the object, at the object's left, or at the object's right, below the object, or above the object.
11. *Relative position: object – surrounding space* — This category refers to information about the object's position within its surrounding space (i.e., on the ground or not, and if not, where exactly in the air).
12. *Size: agent* — This category refers to information about the size of the agent (size of the entity as a whole).
13. *Size: object* — This category refers to information about the size of the object (as in the recipient of an action) as a whole.
14. *Size: instrument* — This category refers to information about the size of the instrument as a whole.
15. *Shape: agent* — This category refers to information about the shape of the agent as a whole.
16. *Shape: object* — This category refers to information about the shape of the object (as in the recipient of an action) as a whole.
17. *Shape: instrument* — This category refers to information about the shape of the instrument as a whole.
18. *Shape of a part: agent* — This category refers to the shape of a part of the agent (shape other than usual shape, or shape information that is not included in information about the overall shape).

19. *Shape of a part: object* — This category refers to the shape of a part of the object (shape other than usual shape, or shape information that is not included in information about the overall shape).
20. *Shape of a part: instrument* — This category refers to the shape of a part of the instrument (shape other than usual shape, or shape information that is not included in information about the overall shape).

Analysis

Six semantic events that were central to the cartoon stories were selected as the focus of the analysis:

Event 1: Jerry (the mouse) is holding a spike, with which he is about to stab Tom (the cat), who is standing in front of him while Jerry faces his back.
Event 2: This refers to the past action of this stabbing with the focus on the recipient, i.e., Tom having been stabbed by Jerry (passive syntax/past tense).
Event 3: Tom is being pinned to the ground by a spike with which he has been stabbed and which is now stuck in his tail.
Event 4: A park keeper has lifted Tom up to carry him away, grabbing him with his hand by the scruff of the neck.
Event 5: A dog is holding Tom in front of him in the air by grabbing him around the throat, strangling him.
Event 6: The dog has lifted Jerry up by the tail and is dangling the mouse with the extended arm in front of his body.

In all, 58 gesture-speech compounds were included in the analysis, of which between 9 and 12 compounds were generated in association with each of the six events.

Analysis of the verbal and gestural material

The *verbal extracts* that have been analyzed here are 'ideational units' (see Butterworth 1975). This means that the length of the speech extracts was determined by how much information a narrator generated with regard to a certain semantic idea, or semantic event, without being interrupted by verbal information that referred to a different semantic idea (for examples, see Holler and Beattie 2002). Furthermore, isolated sentences or small groups of sentences form the basis of the speech extracts in research on discourse ambiguity and inferences (see, for example, Clark and Clark

1977, Singer 1994, Schank 1972). Thus, we assume here that any interlocutor makes inferences at the level of individual sentences, and we attempt to take into account those inferences and to analyze the interaction of gesture and speech with regard to it.

The information of two different but directly successive *iconic gestures* was considered when these were generated while the narrator described the same semantic event verbally and if they occurred in very close temporal relationship, which makes them seem to merge into one. Although iconic gestures are considered as being global-synthetic (see McNeill 1992), an extremely close temporal relationship between two hand movements could be interpreted in such a way that one gesture is not 'sufficient' to represent all the information that the speaker 'intended' to encode gesturally.

A tripartite scheme for coding the verbal information

In the next step of this analysis, the information contained in each extract of speech was scored with regard to 20 different semantic categories (as seen above). The information was scored with three different informational values, namely 0, 0.5, or 1, depending on whether, and how, information concerning the individual semantic features was provided. The criteria for each of the informational values are the following:

0 = The information is not explicitly provided and possible inferences that may be made by an interlocutor would be random and rather weak, or the number of possible inferences is relatively unlimited, which thus allows for more variability among interlocutors in the interpretation of information (e.g., when the maximum information about the exact point on an object, which is contacted by a movement, is the identity of this object, then this still includes an almost unlimited number of potential inferences about where exactly the recipient's body might have been touched).

0.5 = The information is provided implicitly in the sense that the actual words are not contained, but those words that are contained in the utterance allow strong implications in that most interlocutors would make identical inferences, or alternatively the number of possible inferences is relatively limited, which means that the interpretations of different interlocutors will be rather similar (e.g., when the maximum information about the size of the agent is 'a human', then the possibilities range roughly between the size of a child and the size of a grown up).

1 = The information is explicitly provided, i.e., the information is contained in the words that are directly mentioned, or clearly implied due to a precise underlying definition (the words 'to strangle', for example, clearly imply that the point of contact is the recipient's throat).

Here, we provide a number of examples to illustrate our judgments regarding the informational values 0, 0.5, and 1.

Concerning the direction of a movement, for example, information has been scored with 0.5 when the verb 'to pick up' was used. A common inference here would probably be that something is being picked up from the ground. However, other possibilities are that something is being picked up from a shelf next to the agent (demanding a sideward movement), from above the agent (demanding downward movement), or in some cases even that someone is picking someone else up, in the car on the way to work, for example. The verb 'to lift up', on the other hand, leaves no ambiguity about the interpretation with respect to direction, because it refers exclusively to upward movement.

Concerning the semantic category 'action: force', none of the verbal utterances analyzed contained explicit information about the force that is associated with the action, which would be contained in utterances such as 'he's stabbing him *hard*'. However, in our sample the expression 'to clench' did occur, which, according to the *Oxford Dictionary of Current English,* is defined as 'to close (the teeth or fingers) tightly, to grasp firmly'. Thus, in this case, the information was scored with 1 with respect to the information about force associated with the action, because the definition says clearly that the action is accomplished with force. The verb 'to strangle', on the other hand, is defined as 'to squeeze the windpipe or neck of, esp. as to kill' (*Oxford Dictionary of Current English*). This expression is thus, according to the dictionary, not clearly defined in terms of the force that is associated with the agent accomplishing the strangling, because it depends considerably on the agent's strength. In everyday life, the strength of two opponents would probably be assumed as being relatively equal (except when an adult strangles a small child). With regard to the semantics of a cartoon world, however, 'strangling' is equally likely to be accomplished without much force, because cartoon characters often possess extraordinary strength while acting upon a considerably weaker counterpart. In short, if someone clenches the hand, this refers to a forceful act regardless of the agent's strength, whereas the amount of force that is associated with the act of strangling depends very much on the strength of the agent, especially considering that we are talking about a cartoon world. This difference has been captured by scoring 'to strangle' with 0.5 (because it is usually assumed that it is being carried out in a forceful manner) instead of with 1 regarding the category 'action: force'. Furthermore, the expression 'to stab' has been scored with 0.5 in this respect, because an interlocutor is likely to assume strongly that the stabbing is accomplished in a forceful manner, although this is not necessarily the case (everyone knows of film scenes in which one person stands closely in front of another person who suddenly collapses because the other person slowly slides a knife into his opponent's body [and this involves less force than striking out far and moving the knife quickly into someone's body]). Other expressions, such as

'to poke', 'to prod', 'to stick into', or 'to put through' have been scored with 0 concerning the amount of information that they provide about the force associated with the action, because all these actions could equally likely have been accomplished with any degree of force so that common associations are rather likely to be weak or nonexistent.

Another example refers to the semantic category 'action: body-parts involved'. The words 'he's got a stick', for example, could mean that the agent has a stick in his pocket, in his hand, or somewhere at home. A number of these possible interpretations are not associated with any body-parts at all (i.e., those that refer to generally possessing an object [having it at home]), and the second interpretation in terms of 'having the stick right there' could be associated with different body-parts or no body-parts at all (in the pocket). Thus, the information 'to have got something' has been scored with 0 with respect to information about the body-parts that are used to accomplish the action, because associations regarding this semantic aspect can be considered as either nonexistent or very weak. The words 'to hold something', on the other hand, imply clearly that an action is happening, and thus that body-parts must be involved somehow, be it directly or indirectly (e.g., by operating some kind of machinery). Although in the context of a cartoon story actions often are accomplished using the teeth or the feet, a strong association here is that the hands accomplish the action directly. Consequently, verbs such as 'to hold' have been scored with 0.5 regarding the information that the verbs provide about the body-parts that are being used to accomplish the action.

With regard to the semantic category 'entity: object', one utterance was considered as representing information that an object is an integral part of the action only implicitly. The narrator said 'Jerry's gonna put this cue through his tail', which is most likely to be interpreted in the way that the agent is putting the cue through another character's tail, but it could also be interpreted in the sense that the agent sticks the instrument through his own tail (which also does not seem so unlikely considering that the semantic context is a cartoon story). Thus, the information has been scored with the informational value 0.5 with respect to the category 'entity: object'. However, the sentence 'he's stabbing him in his tail', which was the format that most of the other narrators used to describe the same event, has been considered as explicit in this respect because it contains a grammatical object that specifies that another character must be involved (him vs. himself), and was thus scored with 1 regarding the same semantic category.

Concerning the semantic category 'relative position: object-instrument', the statement 'he's been stabbed in his tail' has been scored with 0.5, because it is not entirely clear if the instrument is behind, in front of, or to one side of the object. This is because the tail could be relatively long and

thus could curl toward the front (e.g., as one is usually able to observe with cats when they are sitting down; their tails very rarely point away from their backside in a straight line, but rather curl around and are thus rather at the side or almost in front). However, it is assumed a strong inference that the tail of an animal is behind it. A score of 1, however, has been given for those statements that described the relative position of object and instrument using the words 'he's being stabbed in the bum', for example, because an animal's backside is certainly not at the front or the side.

By mentioning the names 'Tom (the cat)' or 'Jerry (the mouse)' and the words 'bulldog', 'man', or 'cue' (as in billiard cue), a speaker provides explicit information about the size of the entities that are part of the events analyzed, because they are all of a relatively defined size, and thus the informational value 1 was given regarding information about the size of agent, object, or instrument. If, on the contrary, the characters were only mentioned by pronoun ('he' or 'him'), the information has been scored with 0.5 regarding size information, because it is not explicitly stated to which character the speaker is referring. However, due to the number of possibilities for inferences being limited here (the type of cartoon story analyzed here employs a discrete set of characters, usually consisting of Tom and Jerry, and one or two other characters, such as the bulldog or the park keeper), this type of information was scored with 0.5 for the category 'size: agent', 'size: object', or 'size: instrument'. The same applies when a narrator referred to the park keeper as 'a human' or 'a person', and to the instrument as 'a poker' or 'a pole', because an interlocutor will probably make strong inferences concerning the size of these entities (e.g., there is a quite general assumption concerning the size of a poker for a fireplace and a pole that one uses as an instrument to stab someone, although these entities are not defined in size), or the possibilities for plausible inferences are relatively limited (e.g., child or grown-up).

A human, a person, a pole, and a poker (as well as a cat, a mouse, a man, a bulldog, and a cue) are also all defined in their shape, namely human shaped or thin and oblong-shaped (or mouse, cat, or dog shaped). Thus, the score 1 has been given in all these cases concerning the semantic category 'shape: agent', 'shape: object', or 'shape: instrument'.

A slightly different case here represents the word 'spike' used for refer-ring to the instrument that Jerry uses to stab Tom, because it refers to the sharpness of the instrument, i.e., the pointy shape of the ending of the instrument, but not necessarily to the size or the shape of the instrument as a whole. Hence, this kind of information has been scored with 1 regarding the semantic category 'instrument: shape of a part', but with 0 concerning the categories 'size: instrument' and 'shape: instrument'.

Apart from the basic scoring criteria concerning how explicit or implicit certain information is, some additional rules were set up. First, according

to the definition of the semantic category 'shape of a part (of the agent, object or instrument)', information was scored additionally to when information about the entity as a whole had already been scored, only when the shape of a part was shaped in a special way (e.g., the hand was clenched into a fist). Furthermore, to avoid double scoring, information was not scored when it was automatically defined by the shape of the entity as a whole (e.g., 'the pointy needle'; here the name of the instrument already implies that the instrument is pointy). However, when the information was additional, i.e., the information was *not* clearly defined by the information about the shape of the entity as a whole, the information was scored as 'the shape of a part of the instrument' (e.g., 'with a pen or something that he's got, with a sharp instrument'. Because something like a pen can indeed be relatively blunt, the information that the instrument is sharp can be considered as additional and as lessening the degree to which an interlocutor would have to make inferences about the shape of the instrument's ending).

Furthermore, when a narrator used the passive voice and did not mention the agent (e.g., 'he's been stabbed' [left out: by the mouse]), then a score of 0.5 instead of 1 was given, because an interlocutor is likely to infer that an agent must somehow be involved in the action, but it is not explicitly stated (semantic category 'entity: agent' = 0.5 [informational value]). Moreover, because the passive mode shifts an interlocutor's attention toward the recipient of an action (e.g., Tom, who has been stabbed) and away from the agent, it can be assumed that any interlocutor makes considerably less strong inferences concerning semantic features that relate to the agent. Consequently, such cases, in which it was assumed that an interlocutor would not make strong inferences about information relating to the agent, were scored with 0 instead 0.5. For example, an interlocutor might normally assume strongly that the agent is holding an object with the hands (e.g., 'he's holding him by the scruff of the neck'), whereas a sentence in the passive mode (e.g., 'he's being held by the scruff of the neck) might evoke a mental representation of a scene, which an agent is an integral part of to a much lesser extent or not at all. The semantic categories 'action: body-parts' and 'shape of a part: agent' were hence scored with 0 instead of 0.5. Furthermore, information about the size or the shape of the agent is obviously not provided at all, so these semantic categories were also scored with 0.

We will now use a larger extract of speech, taken from the present database, to exemplify the process of scoring for the entire range of semantic categories applied here:

'Jerry's about to . . . strike him with something quite sharp, maybe in his behind. Looks like it is something with like a narrow point'

Table 2. *Example of the scoring procedure for verbal material as applied in the present study*

Semantic category	Informational value
Entity 1 — agent:	**1**

By mentioning that '*Jerry*' is '*doing* something', the narrator provides explicit information that an agent is an integral part of the event.

Entity 2 — object:	**1**

Because the utterance contains the pronoun 'him', which takes the position of a grammatical object, explicit information that an object, in terms of the recipient of an action, is an integral part of the event is provided.

Entity 3 — instrument:	**1**

The utterance contains explicit information that an instrument, in terms of a third entity, is an integral part of the action, because it contains the words 'to strike *with something* sharp'.

Action: body-parts used	**0.5**

Which body-part the agent uses to accomplish the action is not explicitly mentioned, but the possibilities for inferences are limited. In the context of a cartoon story, it is not uncommon that body-parts such as the feet or the teeth may be used to accomplish an action as opposed to the hands, but the hands can still be considered as most strongly associated with the action of holding an instrument.

Action: direction of the movement	**0**

The utterance contains neither explicit nor implicit information about the direction of the movement, because it is not defined whether the agent moves the instrument upward, downward, forward, etc.

Action: point of contact defined	**1**

The speaker provides explicit information about the point that is contacted by the movement, namely the recipient's behind (it is to note that the narrator's uncertainty about this, expressed by the word 'maybe', is due to the ambiguity in the cartoon drawing and is therefore not being considered here).

Action: force	**0**

The utterance contains no explicit or strong implicit information about the degree of force with which the action is accomplished, because the act of striking an object could refer to a forceful act or it could be accomplished very easily, without much effort being required from the agent's side (e.g., if the agent is very strong, or the strike is weak).

Relative position: agent — object	**0**

The extract of speech contains no explicit information about the relative position of agent and object. Although one could argue that an interlocutor might assume that the action is accomplished in front of the body, and that therefore the agent might have turned toward the object so that he or she is facing the object, this inference has been considered as possible but not as strongly implied. In contrast to inferences concerning the body-parts involved in the action, which has been considered as a rather strong association, the number of possibilities for how agent and object are positioned relative to each other is relatively *less* limited (e.g., next to each other, in front, behind, one above the other in a tree, etc.).

Table 2.   *(continued)*

Relative position: agent — instrument   **0**

As mentioned in association with the semantic category 'relative position: A – O', the possibilities concerning the relative position of agent and instrument are rather manifold. The agent could be holding the instrument at the side of the body, in front of the body, above the head, etc. Thus, the utterance has *not* been considered as implying strongly how agent and instrument are positioned relative to each other.

Relative position: object — instrument   **1**

The speech extract contains explicit information about the position of the instrument relative to the object, namely on the object's behind.

Relative position: object — surrounding space   **0**

The speaker does not provide any explicit information about where the object is positioned relative to its surrounding space. With regard to a cartoon world, it is not necessarily most likely that a character is standing on the ground (quite a large proportion of actions take place in mid-air or on some kind of surface other than the ground). Thus, the utterance was considered as not providing information that strongly implies a certain relative position of the object in its surrounding space.

Size: agent   **1**

Because the present analysis has been conducted taking into account the rough context in which the speech extracts occurred, i.e., a Tom & Jerry cartoon story, the name 'Jerry' identifies a certain cartoon character (a mouse) and thus provides explicit information about the size of this character.

Size: object   **0.5**

The object, i.e., the recipient of the action, is being referred to with only the pronoun, but the speech extract does not explicitly determine the identity to whom this pronoun refers. Because the context is a 'Tom & Jerry' cartoon story, the possibilities for inferences are limited; it is likely that the pronoun refers to 'Tom'. However, there can be no absolute certainty here, because other characters are getting involved from time to time.

Size: instrument   **0**

The only information provided that relates to the instrument is that it is sharp and has a narrow point. Based on this information, however, no assumptions about the size of this instrument as a whole can be made.

Shape: agent   **1**

As mentioned in association with information about the size of the agent, the extract of speech can be considered as providing explicit information about the shape of the agent, because 'Jerry' is known as being a mouse, and thus of a relatively defined shape.

Shape: object   **0.5**

Information about the shape of the object, i.e., the recipient of the action, is, as concerning its size, only implicitly provided. This is because the pronoun 'him' is being used, which does not specify the identity of the character. However, it can also be argued that the possibilities of the identity of the character other than Jerry are limited in this context, but not exclusive, because characters other than 'Tom' and 'Jerry' sometimes play a part in the action.

Table 2. *(continued)*

Shape: instrument   **0**

The speech extract does not contain any explicit information about the shape of the instrument as a whole, nor does it contain information that strongly suggests a certain shape of the instrument as a whole. A relatively 'sharp, narrow point' could be, for example, the corner of a book as well as the tip of a pencil.

Shape of a part: agent   **0.5**

Although the verbal utterance does not contain any explicit information about how a certain part of the agent's body is shaped apart from its normal shape, it has been considered as containing implicit information. For the semantic category 'action: body-parts involved', we argued that the action of 'striking someone with something' implies that it is the hands with which the instrument is being held. Thus, one would also infer that the hand is holding the instrument by gripping around it, which is the most likely way of holding onto an object.

Shape of a part: object   **0**

Information about the shape of any of the object's body-parts is not provided by the extract of speech, neither explicitly nor implicitly. This is because, in this case, none of the object's body-parts are shaped in a special way (e.g., a fist), which would not be automatically defined by the shape of the object as a whole.

Shape of a part: instrument   **1**

Information about the shape of a part of the instrument is explicitly provided here, namely by mentioning that the instrument is sharp and has a narrow point.

A binary coding scheme for iconic gestures

The coding scheme for iconic gesture is a binary instead of a tripartite scheme because, based on the criteria for scoring applied here, none of the iconic gestures analyzed were considered as representing certain information in a way that would have made a value other than 0 or 1 necessary. In all the cases, information was either represented in the gesture or not represented. Of course, theoretically it could be that in some cases it is less easy to decide whether or not certain semantic information is represented by an iconic gesture, which might then be considered as an informational value 0.5. However, this was not the case regarding the iconic gestures analyzed in the present study. (The only semantic category that showed to be slightly harder to decide about was 'action: force', because it is a more subjective judgment of how forceful or weak a movement is as compared to, for example, a judgment concerning the direction of the movement. However, there were no strong disagreements [see the values for Cohen's Kappa at the end of this section], and a third informational value would not have eliminated the subjectivity of the judgment.) The gestural information was thus scored with 0 or 1 according to the following criteria:

0 = The information that is contained in the accompanying speech or in the original scene is represented by the iconic gesture.
1 = The information that is contained in the accompanying speech or in the original scene is not represented by the iconic gesture.

Some examples of iconic gestures are described subsequently to illustrate our judgments for the informational values 0 and 1. For example, most narrators referred gesturally to 'event 2' by clenching the hand into a fist to imitate that he or she is holding the instrument, and by moving the arm from near the body downward/forward to imitate the agent stabbing the object in front of the agent. A gesture of this kind has been scored with 1 concerning the semantic category 'action: direction of the movement'. (The reason that some gestures seemed to be aiming at an object more frontally and some more downward lies here in the ambiguity of the drawing of the cartoon, which is ambiguous in that Jerry is aiming at Tom, but the spike is not quite touching him yet. Thus, according to the cartoon, the movement could continue forward as well as downward.) Another narrator, however, referred to 'event 2' by saying 'he's been spiked with that poker thing', while his or her left hand moves from the right side into the center of the narrator's gesture space (the arms were hanging crossed over the legs in rest position before) and the hand clenches into a fist; it remains in front of the narrator's body for a moment, and then the hand moves out of the gesture space and into rest position at the left side of the narrator. This gesture represents the agent clenching the hand around the instrument and holding it in front of the body, but it does not represent any information about the agent stabbing the object in front. Thus, the gesture here has been scored with 0 concerning the semantic category 'action: direction of the movement', because the gesture does represent some movement, but this is not 'critical' movement with regard to the original scene (i.e., it does not represent the direction of the actual movement suggested by the cartoon).

Furthermore, referring to 'event 1–3', most narrators provided the information that an object, which the agent is acting upon, is a part of the action by stopping the movement of the arm (representing the act of stabbing) relatively abruptly. This gesture has hence been scored with 1 regarding the category 'entity: object'. As opposed to a movement that slowly loses swing, this abrupt stopping signifies that some kind of entity seems to present a resistance that hinders the arm from moving further. However, when a narrator did not stop the movement of the arm abruptly, but continued the movement until it lost swing, this was scored with 0 for the semantic category 'entity: object'.

Finally, one narrator represented 'event 3' gesturally by extending the index finger of the right hand, moving the hand upward from the height of the seat to about the height of the shoulder in a straight line and then down again, and then the hand moved behind the narrator and backward away from the narrator at the height of the seat, the index finger still extended. This iconic gesture represents information about the size and the shape of the instrument (thin and oblong-shaped and about three feet high) so that it was scored with 1 regarding both semantic categories, 'size: instrument' and 'shape: instrument'. However, from a gesture, which involves that the narrator imitates holding an instrument in the clenched hand, information about the size and the shape of the instrument as a whole is not inferable. The only information that is provided here by the gesture is that the instrument is thin and oblong-shaped to the extent to which the hand imitates grasping the instrument. Such a gesture was thus scored with 0 regarding information about the shape of the instrument as a whole (but with 1 for the semantic category 'shape of a part').

One example of an iconic gesture will now be provided to illustrate the process of scoring for the entire range of semantic categories. The following gesture occurred together with the extract of speech that has been analyzed above:

'Jerry's about to . . . strike him with something quite sharp, maybe in his behind. Looks like it is something with like [a narrow point]'
[right hand rises to shoulder height, the back of the hand is pointing upward/backward, the fingertips are pointing forward and all touch each other, the hand moves quickly forward with the fingertips ahead, the movement stops abruptly and the hand moves immediately back again (2x)]

The inter-observer reliability for the entire range of scoring criteria has been tested using a naïve judge, who was given an example of the scoring procedure for one gesture-speech compound with the reasoning for those decisions, and who then judged the kind of implicit and explicit information that he gained from certain speech extracts and iconic gestures (the judge scored four speech extracts that were randomly chosen for each semantic event, i.e., 24 speech extracts and iconic gestures out of a total of 58 [= 41%]). Cohen's Kappa, a coefficient of agreement, was calculated for all 20 semantic categories together (due to the otherwise small number of data points in the individual cells) and resulted for speech in K = .96 and for iconic gesture in K = .97. The percentage agreements were calculated for the individual semantic categories, which, for speech, ranged from 91.7 percent for 'size: instrument', 'shape: instrument', and 'shape of a part: instrument' to 100 percent for 'entity: agent', 'entity: object', 'entity:

Table 3.   *Example of the scoring procedure for gestural material as applied in the present study*

Semantic category	Informational value
Entity 1 — agent:	1

Because the narrator transfers here into the role of the agent to perform the action, the gesture does provide the information that an agent is an integral part of the action.

Entity 2 — object:	1

Through the relatively abrupt stopping, the information is provided that some kind of entity must be preventing the hand from moving further, that is another entity that the agent is acting upon, i.e., an object.

Entity 3 — instrument:	1

Information that an instrument, in terms of a third entity, is an integral part of the action is provided in that the narrator imitates holding something thin and oblong-shaped in the hand.

Action: body-parts involved in the action	1

Information about the body-parts that the agent uses to accomplish the action is provided in that the narrator re-enacts the action using the hands, as the agent does in the cartoon drawing.

Action: direction of the movement	1

The forward movement of the gesture is highly similar to the movement that the cartoon suggests.

Action: point of contact defined	0

Although the iconic gesture represents that an object that the agent is acting upon is part of the action, it does not provide any information about the identity of this entity, and does thus not provide any information about the exact point of contact with regard to the object's body.

Action: force	0

The gesture has been considered as providing no information about a particular degree of force that is associated with the action, because it does not involve that the arm is being moved especially powerfully or weakly toward the imaginable object.

Relative position: agent — object	1

Information about the relative position of agent and object is represented in that the narrator shows to be acting upon an object in front of him or her by moving the hand with the imaginable instrument forward.

Relative position: agent — instrument	1

The gesture provides also information about the relative position of agent and instrument, because the narrator imitates holding the instrument at about shoulder height at the side of the body.

Relative position: object — instrument	0

Information about the relative position of object and instrument is not provided here, because the gesture does not represent any information about the object's identity and thus not in which direction the object is facing. Thus, the agent could be stabbing the object in the face, in the right side, the left side, in the back, etc.

Table 3.   *(continued)*

Relative position: object — surrounding space   **0**

According to McNeill, the space in which iconic gestures are generated 'is not the speaker's space, but a fictional space, a narrative space that exists only in the imaginary world of the discourse' (1992: 1). Thus, iconic gestures are by definition not suitable to represent information about where a certain entity is positioned relative to its surrounding in the 'real' world.

Size: agent   **0**

Although the narrator transfers here into the role of the character, he or she does not adopt the character's size.

Size: object   **0**

Because the narrator does not represent any information about the identity of the object, the gesture does also not represent any information about the object's shape.

Size: instrument   **0**

Information about the size of the instrument as a whole is not represented by the iconic gesture, because the narrator imitates that the instrument is being held in his or her hand, but does not represent the instrument as a whole.

Shape: agent   **0**

As mentioned in association with the category 'size: agent', the gesture does not represent any information about the shape of the agent, because the agent is represented by the narrator transferring into the role of the agent, but he or she does not adopt the agent's shape.

Shape: object   **0**

The iconic gesture does not represent any information about the object apart from the fact that an entity, in terms of the recipient of the action, is part of the semantic event, and thus it provides no information about the shape of the object.

Shape: instrument   **0**

Information about the instrument is provided by the gesture only in that the narrator is holding an instrument in the hand, which again provides information about the shape of a part of the instrument, but not about the shape of the instrument as a whole.

Shape of a part: agent   **1**

Because the narrator transfers into the role of the agent and re-enacts the action that the agent performs, the gesture provides the information that the agent's hand is shaped in a way it is when gripping something thin and oblong-shaped.

Shape of a part: object   **0**

Again, because the gesture here does not represent any information about the object other than the fact that an object is an integral part of the semantic event, no information is provided concerning the shape of any of the object's parts.

Shape of a part: instrument   **1**

The gesture does represent information about the shape of a part of the instrument, which the narrator imitates holding in the hand. This way, it becomes clear that the object must be at least in part oblong-shaped and so thin that it is possible to grasp it with the hand.

---

instrument', 'action: body-parts', 'action: direction of movement', 'action: point of contact defined', 'relative position: agent-object', 'relative position: agent-instrument', 'relative position: object-instrument', 'size: agent', 'size: object', 'shape: agent', 'shape: object', 'shape of a part: agent', and 'shape of a part: object'. For iconic gesture, the percentage agreements ranged from 87.5 percent for 'action: force' to 100 percent for 'entity: agent', 'entity: instrument', 'action: body-parts involved', 'action: direction of movement', 'action: point of contact defined', 'relative position: agent-object', 'relative position: agent-instrument', 'relative position: object-instrument', 'relative position: object-space', 'size: agent', 'size: object', 'size: instrument', 'shape: agent', 'shape: object', 'shape: instrument', 'shape of a part: agent', and 'shape of a part: instrument'.

Results

Using the criteria for scoring that have been described in the 'Method' section, 58 gesture-speech compounds were analyzed to find out how information is represented by iconic gesture and speech according to the 20 semantic categories applied in this study. In Holler and Beattie (2002), the same basic gesture-speech compounds were analyzed, but only with regard to position and size information. Furthermore, as a subsequent step, their first analysis involved that the scores for each informational value were then summarized for each of the semantic categories, for gesture and speech separately, across all the gestures and all the speech extracts. This procedure is exemplified in Table 4. (Note that Table 4 shows the data analyzed according to the revised criteria for scoring the verbal and gestural material as applied in the present study, which means that the data look somewhat different from those in Holler and Beattie [2002].)

This type of data analysis was regarded as providing a relatively good indicator of how certain information is represented by gesture and speech, because the original data (i.e., before the scoring criteria were modified, and when only two of the original three types of informational values were considered) was very clear-cut with the data being split almost exclusively between the informational values 0 and 1 only. However, because the present data shows a more complex pattern for most of the semantic categories and because we aim to advance our previous research, aggregate data of this type was considered here as not offering sufficient insight to draw plausible conclusions concerning the interaction of gesture and speech. Consequently, the rather complex database has been dismantled according to the six different possible combinations of informational

Table 4. *Overview of the distribution of frequencies across the individual semantic categories for each of the informational values, for gesture and speech separately*

Semantic category	Gesture		Speech		
	Informational value				
	0	1	0	0.5	1
1. Entity: agent	6	52	8	15	35
2. Entity: object	5	53	0	1	57
3. Entity: instrument	0	27	0	3	24
4. Action: body-parts	6	52	22	33	3
5. Action: direction o. movem.	23	35	44	11	3
6. Action: point of contact	53	5	24	0	34
7. Action: force	49	9	44	13	1
8. Rel. position: agent-object	12	46	57	0	1
9. Rel. position: agent-instr.	5	22	27	0	0
10. Rel. position: object-instr.	26	1	12	13	2
11 Rel. position: object-space	58	0	52	1	5
12. Size: agent	58	0	23	15	20
13. Size: object	58	0	0	45	13
14. Size: instrument	26	1	17	5	5
15. Shape: agent	58	0	23	12	23
16. Shape: object	58	0	0	45	13
17. Shape: instrument	25	2	13	0	14
18. Shape of a part: agent	6	52	24	32	2
19.Shape of a part: object	26	32	58	0	0
20. Shape of a part: instrument	2	25	12	8	7

values (see Table 5). Thus, rather than describing by which type of infor-mational value a certain semantic feature is most frequently represented in one of the communicational channels *irrespective of the other,* the present analysis allows for an exact determination of how often *combinations* of informational values occurred. Thus, the present analysis describes directly how often gesture and speech interact in a particular kind of pattern.

The advantage of this way of analyzing the data becomes clear when we compare, for example, the distribution of frequencies for the semantic cat-egory 'force' shown in Table 4 to those shown in Table 5. The frequencies that are displayed in Table 4 may seem to suggest that the 49 cases in which the gestures did not provide any information are to a large extent con-gruent with the 44 cases in which no verbal information was represented. Furthermore, one may assume that the 9 cases in which the gestures did represent information about the degree of force that is associated with a certain action, are, apart from a few cases, congruent with the 13 cases

Table 5.   *Overview about the distribution of frequencies across all six combinations of informational values for each semantic category*

Semantic category	Combination of informational values					
	G – S 0–0	G – S 0–0.5	G – S 0–1	G – S 1–0	G – S 1–0.5	G – S 1–1
1. Entity: agent	3	2	1	5	13	34
2. Entity: object	0	0	5	0	1	52
3. Entity: instrument	0	0	0	0	3	24
4. Action: body-parts involved	5	1	0	17	32	3
5. Action: direction of move.	19	4	0	25	7	3
6. Action: point of cont.	24	0	29	0	0	5
7. Action: force	38	10	1	6	3	0
8. Rel. position: agent-object	12	0	0	45	0	1
9. Rel. position: agent-instr.	5	0	0	22	0	0
10. Rel. position: object-instr.	12	12	2	0	1	0
11 Rel. position: object-space	52	1	5	0	0	0
12. Size: agent	23	15	20	0	0	0
13. Size: object	0	45	13	0	0	0
14. Size: instrument	16	5	5	1	0	0
15. Shape: agent	23	12	23	0	0	0
16. Shape: object	0	45	13	0	0	0
17. Shape: instrument	12	0	13	1	0	1
18. Shape of a part: agent	5	1	0	19	31	2
19. Shape of a part: object	26	0	0	32	0	0
20. Shape of a part: instrum.	1	1	0	11	7	7

* G = gesture; S = speech

in which speech represented implicit information. However, a look at the frequencies shown in Table 5 reveals that this is not so. Rather, in a considerable number of cases, speech represented implicit information about the semantic feature 'force', but the gesture represented no information about this semantic feature, whereas in a number of other cases the implicit verbal representation of the semantic feature 'force' *was* accompanied by the gestural representation of force.

Table 5 reveals a number of diverse patterns as to how the gestural and the linguistic systems interact to represent semantic information, and none of the semantic features shows an identical pattern of representation. This means that if we consider how the frequencies are distributed across the six possible combinations of informational values regarding each individual semantic category, the distribution of frequencies is rather homogenous for some of them, whereas it is rather heterogeneous for others.

At the same time, however, certain similarities in the interaction patterns can be found concerning some of the semantic categories. In order to

Table 6. *Illustration of the classification procedure applied in the present study*

G	S	Entity: agent	Entity: object	Entity: instrum.	Action: body-p	Action: direction	Action: p. o. c.	...
0	0	5.17%	0.00%	0.00%	8.62%	32.76%	41.38%	...
0	0.5	3.45%	0.00%	0.00%	1.72%	6.90%	0.00%	...
0	1	1.72%	8.62%	0.00%	0.00%	0.00%	50.00%	...
1	0	8.62%	0.00%	0.00%	29.31%	43.10%	...	...
1	0.5	22.41%	1.72%	11.11%	55.17%	12.07%	...	...
1	1	58.62%	89.66%	88.89%	5.17%	5.17%	...	...

capture and summarize these similarities, the following steps were undertaken: First, the percentages were calculated for each cell (see Table 6), taking into account the different numbers of gesture-speech compounds that have been analyzed with regard to each semantic category, i.e., 27 or 58.

A second step involved an attempt to simplify the rather complex data set using a kind of data reduction technique. The intention was to classify the semantic categories in groups according to the degree of similarity in the underlying pattern of representation. The basic idea was that the frequency for each individual cell is 1/6, or 16.67 percent, if the data is equally distributed across the six different combinations of informational values. However, if we assume the other extreme, that the frequencies were distributed in a highly heterogeneous manner across the six combinations of informational values, then we could expect that a certain combination of informational values prevails. A prevalent pattern of interaction of both communicational channels could then be considered as characterizing how information about a certain semantic feature is represented by gesture and speech. By comparing the distribution of frequencies that the present data show to the frequencies that would result when the data were distributed homogeneously, we apply the basic principle that underlies a χ^2 Test, namely the comparison of expected and observed frequencies.

In order to classify the semantic categories as represented primarily by one of the six interaction patterns, the following criteria were set up: First, a certain semantic feature had to be represented by at least 50 percent (that is 16.67 percent, the frequency for each cell assuming equal distribution, times three) or more of the gesture-speech compounds by one of the interaction patterns, and second, the next highest value had to be at least 10 percent less than the highest value (see gray shaded cells in Table 6 for examples). These principles were established to ensure that a considerable amount of gesture-speech compounds represented information in the same way (referring to the combination of informational values in both

communicational channels), and that this amount differed substantially from the next most frequent pattern of interaction.

According to the data reduction technique applied here, the 20 semantic features can be classified into different groups, namely those that are predominantly represented by iconic gestures but not by speech and vice versa (1–0; 0–1), those that are represented by both communicational channels (1–1), those that are predominantly represented by neither of the two channels (0–0), those that are predominantly implicitly represented by speech and at the same time gesturally (0.5–1), and those that are predominantly implicitly represented by speech but not by gesture (0.5–0).

Figure 1 shows how the semantic features summarize into the different kinds of groups. To interpret the figure, it might be helpful to imagine the two axes as representing gesture and speech (as specified in Figure 1), and thus how both communicational channels interact predominantly to represent the individual semantic features according to how often the individual combinations of informational values occurred.

Additionally, certain semantic features can be grouped together, which, according to the criteria applied here, are not predominantly represented in one of these ways, but whose frequencies distribute more
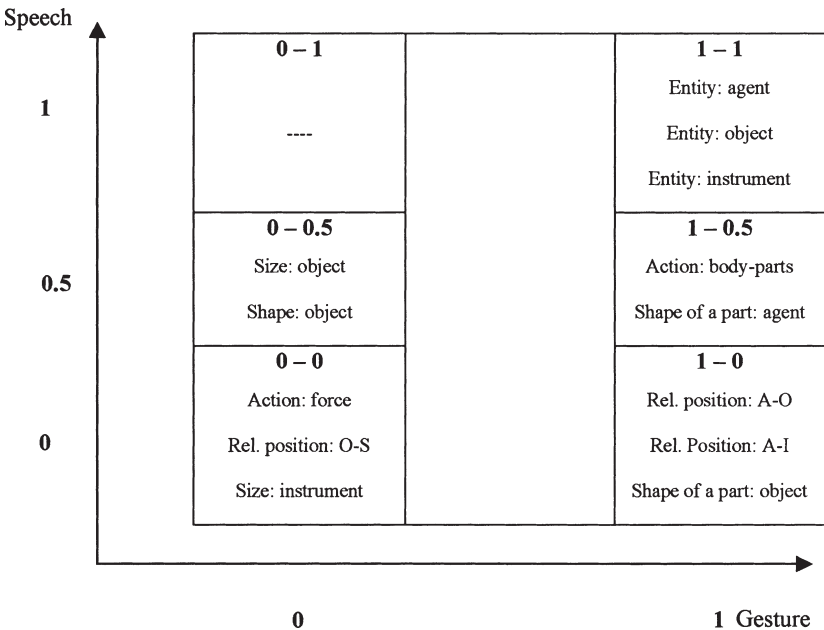

Figure 1.   *Overview of the different groups of semantic features*

homogeneously across all six types of interaction patterns. Those semantic features that did not show a predominant pattern of representation according to these criteria could be summarized based on lowered criteria (however, the representation of these semantic features has to be considered as being less clearly characterized by one predominating interaction pattern, such that they are not displayed in Figure 1). The lowered criteria involved that a semantic feature was represented by a certain interaction pattern in 33.33 percent of the cases (= 16.67 percent times two) or more, plus the next highest value must at least be 10 percent lower than the highest. According to these criteria, the semantic features 'action: direction' (43.10 percent) and 'shape of a part: instrument' (40.75 percent) were both predominantly represented by iconic gesture but not by speech (1–0).

Most of the remaining semantic categories showed, according to the '33.33 percent + 10 percent distance criterion', two combinations of informational values that characterized the interaction of gesture and speech best. Information relating to the category 'action: point of contact' has predominantly been represented neither by gesture nor by speech (0–0; 50 percent), or explicitly by speech but not by iconic gesture (1–0; 41.38 percent). Information about the relative position of object and instrument has predominantly been represented implicitly by speech but not by gesture (0.5–0; 44.44 percent), and also neither by speech nor by iconic gesture (0–0; 44.44 percent). The semantic category 'shape: instrument' was predominantly represented explicitly by speech but not by gesture (1–0; 48.15 percent), or neither by speech nor by gesture (0–0; 44.44 percent).

Only two of the semantic categories were about equally well characterized by three of the combinations of informational values (according to the '33.33 percent + 10 percent criterion'). Information about the 'size of the agent' as well as about the 'shape of the agent' was predominantly represented neither by speech nor by gesture (0–0; 39.66 percent for both semantic categories), as well as explicitly by speech but not by gesture (1–0; 34.48 percent and 39.66 percent), and third, implicitly by speech but not by iconic gesture (0.5–0; 25.86 percent and 20.69 percent).

In conclusion, our analysis categorizes the range of semantic features into different groups according to their predominant pattern of how gesture and speech interact together. Of course, the criteria for the classification that have been applied are somewhat arbitrary. However, we could argue that they do provide a good summary of how gesture and speech operate together to represent certain semantic information. Furthermore, the technique applied summarizes very complex data and offers a way of outlining which of the semantic features are represented by gesture and speech in a similar way, and which in a different way.

Discussion and further analysis

The present study investigated how iconic gestures and speech interact in representing semantic information. We applied a range of 20 semantic categories and a mathematical scheme for coding the information represented by both communicational channels. Thus, the present study clearly goes beyond the study by Holler and Beattie (2002), who investigated how the gestural and the linguistic systems interact while focusing only on the seven sub-categories that relate to the two basic semantic categories 'relative position' and 'size' (see Table 1). Furthermore, in the first study, Holler and Beattie used the same procedure for scoring the information represented in the gestural as well as in the verbal channel, namely both were scored in the absence of the other channel. As pointed out in the introduction, however, this may lead to problems interpreting certain gestural representations apart from relative position, which is the only semantic feature that seems to be communicated by iconic gestures in the absence of speech, at least in a consistent fashion according to Beattie and Shovelton's studies (1999b, 2002). Furthermore, the scoring procedure applied in the present study takes into account *inferences* that interlocutors are likely to make while listening to spoken discourse, and we differentiated here between implicit and explicit information. This was not considered in the first Holler and Beattie study.

McNeill's theoretical position seems to suggest that iconic gestures reveal different semantic aspects about a scene than does speech, and that consequently the gesture is necessary to provide a complete insight into the scene that a narrator has in mind. We would argue, however, that this description of how the verbal and the gestural systems interact to represent semantic information is currently rather vague and general. The findings of the present study suggest that if one uses a semantic feature approach, it becomes obvious that the pattern of how gesture and speech operate together varies considerably from feature to feature. Some semantic features were primarily represented by gesture but not by speech ('relative position: A–O', 'relative position: A–I', and 'shape of a part: object'). Others were primarily represented by speech (but implicitly) but not by iconic gesture ('size: object' and 'shape: object'). Both of these types of interaction pattern support McNeill's argument that speech and gesture provide information about *different* semantic aspects of the same cognitive representation. Other semantic features, however, were predominantly represented by speech (implicitly) and at the same time by iconic gesture ('action: body-parts' and 'shape of a part: object'), and others were represented by both iconic gesture *and* speech (but explicitly) ('entity: agent', 'entity: object', and 'entity: instrument'). Finally, some semantic features

were hardly represented by either speech or gesture ('action: force', 'relative position: O–S', and 'size: instrument'). Thus, the gestural and the verbal information seems to be rather redundant with respect to each other sometimes, which contrasts with McNeill's theoretical position, because it means that iconic gestures do not seem to be strictly integral to the representation of semantic information with regard to the range of semantic features that constitute a certain scene. In conclusion, we can say that the interaction of the gestural and the linguistic systems is more multifaceted than as described by McNeill and that iconic gesture and speech do not necessarily represent different aspects of the same scene. Rather, they only do so with respect to certain semantic features.

Of course, the question arises as to why certain semantic features are represented differently to others. One hypothesis would be that those semantic features that are similarly represented bear some kind of common characteristic. Such a hypothesis would suggest that gesture and speech are two systems, one of which is designed to represent information about a certain array of semantic features and the other is designed to represent information about another set of semantic features, and thus that the gestural and the linguistic systems operate together in a rather static and fixed manner. Regarding the semantic features that showed to be represented in a similar way here, however, it is not immediately obvious as to why this is concerning most of the cases (what, for example, have the semantic features 'action: force', 'relative position: object-surrounding space', and 'size: instrument' in common?).

An alternative hypothesis might be that the gestural and the linguistic systems interact in a much more *flexible* manner, which is linked to the communicational intent of the speaker. Such a theoretical model would assume that it is not possible to determine which semantic features are represented by iconic gestures and which features are represented by speech. Rather than a static and fixed pattern of partitioning, the hypothesis would be that the gestural system may serve different kinds of communicational functions and that, according to the communicational demands of a certain situation, the kinds of semantic features that iconic gestures are used to represent may actually vary. The six possible combinations of informational values that have been introduced in this study (see Table 5) might offer a hint as to what such functions might be.

For example, with respect to those cases in which speech provides only implicit information (informational value 0.5), which means that an interlocutor must infer certain information to understand the full meaning of what is being communicated, one could argue that the gesture might function in the way that it helps the interlocutor to make the right inferences. In other words, one possible communicational function of iconic gestures might be that they disambiguate speech in the sense that they help

to prevent an interlocutor from making the wrong inferences, and that an interlocutor is able to construct a more accurate mental representation of the scene that his communicational partner is talking about. However, because we consider the speaker's intent as the decisive factor that makes the gestural system a *flexible* resource, it must be assumed that the speaker considers it important to communicate certain information unambiguously in a certain situation. In a pilot study testing for the kinds of inferences that people make based on only verbal information, a number of ideational units that contained implicit information concerning certain semantic aspects (according to the present scoring procedure) were presented to different interlocutors. The interviewees were then asked questions concerning the individual verbal statements, which tapped the respective semantic aspects. The interlocutors' answers showed that, indeed, they had made some inferences that are incorrect with regard to the original scene that the narrators were describing. For example, one of the sentences was 'the dog's got hold of Jerry and is dangling him by the tail' and supposed to tap the semantic feature 'action: body-parts'. Some interlocutors assumed that the dog would be holding Jerry with its mouth instead of with its hand as suggested by the original scene in the cartoon. The iconic gesture that originally accompanied the verbal utterance, however, showed the narrator with its arm extended at about stomach height while the hand performed a pincer-like grip, therefore showing quite clearly that the agent was using the hand to dangle Jerry rather than Jerry dangling from the dog's mouth. This example shows how iconic gestures may help to disambiguate speech by providing 'visual clues' for interpreting the verbal information in exactly the way that the speaker has in mind. As can be seen from the data displayed in Table 5 (see the 'Results' section), however, iconic gestures do not always seem to fulfill this kind of function, because verbally implicitly represented information is not always accompanied by a gestural representation of the same semantic feature. This may be true simply because in these cases it was not the primary communicational intent of the speaker to provide unambiguous information about these semantic features.

   Therefore, it makes sense to discuss other communicational intentions that a speaker might have than the disambiguation of speech and for which he or she might use iconic gestures as a resource that can be utilized in the realization of this intention. For example, another communicational intent might be the emphasis of a certain semantic aspect, and this can be seen in those cases in which certain semantic information was represented by both speech *and* iconic gesture. One of the events chosen for the present analysis (event 4) involved a relatively big and strong park keeper picking up Tom

the cat by the scruff of the neck. Without the experimenter asking a question that stressed the semantic features 'force' (e.g., by asking whether Tom the cat is heavy for the park keeper to carry), most of the narrators referred to event 4 either by clenching the hand into a fist, which is then moved downward a bit and up again to imitate the park keeper grabbing Tom and picking him up, or simply by clenching the hand to imitate the park keeper gripping Tom's neck fur as in the following example:

'he's [being picked up by like the scruff of his neck]'
[right hand is resting on the leg, rises slightly while clenching into a fist, then the hand drops back into rest position]

Referring to this example, neither the extract of speech nor the accompanying iconic gesture were considered as providing the information that the action seems to have been accomplished especially easily due to the park keeper grabbing the cat with his big strong hand (i.e., there was especially little force involved). However, in those cases in which the experimenter asked a question that emphasized the semantic feature 'force' (i.e., whether Tom is heavy for the park keeper to pick up), many narrators used an iconic gesture that did provide information about the force involved in accomplishing the action. This is exemplified by the following gesture-speech compound:

'um. .no, ['cause he's just . . . pulling him up by the scruff of the neck]'
[right hand rises to about the height of the stomach, the arm is slightly extended, the back of the hand is up, the fingers are relaxed and hanging down, but the tip of the index finger and the thumb touch each other, open and touch again a few times while the hand is moving up and down (5x)]

In this example, both the extract of speech and the accompanying iconic gesture were considered as representing information about the force that is associated with accomplishing the action, i.e., that especially little force is required. Linguistically, the word 'just' can imply that something is done easily. Gesturally, the quick repetition of the 'picking up movement' can be interpreted as showing that something is done easily (the contrary would probably be a re-enactment of the action performing the movement extremely slowly to show that the object that is being picked up is heavy). Therefore, considering both examples, it shows that the same event might be communicated differently if the conversational demands place emphasis on a certain semantic aspect of an event, the conversational demands being that the experimenter asked for information relating to a particular semantic feature of the event, namely the semantic feature 'force'. Information about the semantic feature 'force', however, was not

represented by iconic gestures that accompanied the participants' narratives of 'event 4' without the experimenter asking the question. An explanation for this could be that 'force' is a semantic feature that is most characteristic for cartoon stories, because most of the actions are accomplished in a rather forceful manner or, on the contrary, extremely easily. The consequence of this might be that the communicational task of simply describing what one is seeing (as the narrators in this study were instructed) does not encourage to emphasize the force that is associated with certain actions because, in the context of a cartoon story, the fact that the actions involve a particular degree or a particular lack of force may seem self-evident. With regard to the example above, one could thus argue that although the information is provided linguistically (he 'just' pulled him), many narrators used an iconic gesture that, contrary to those gestures that were generated not in response to the respective question but referring to the same semantic event, provided information about the force that is involved in the action.

To conclude, in order to explain how the representation of semantic information is partitioned between iconic gesture and speech, one can develop two general hypotheses. Hypothesis 1, which seems to have directed previous research on iconic gesture and speech (e.g., Beattie and Shovelton 1999a, 1999b, 2002; Holler and Beattie 2002) argues that particular semantic features are represented by iconic gesture, and other particular semantic features by speech. However, the present findings do not seem to support this hypothesis. Rather, they suggest that how semantic information is partitioned between gesture and speech is influenced by a number of certain other factors, including the current conversational demands, and that makes an alternative hypothesis seem perhaps more plausible. This latter hypothesis predicts that speech and gesture do not interact in a static and fixed manner, but that the interaction of both systems is rather flexible. This means that a speaker can draw upon the gestural system as a resource, for example, to disambiguate speech in those cases in which the linguistic information requires an interlocutor to make certain inferences and in which the speaker considers it as important to provide relatively unambiguous information. Or it might be the speaker's intent to emphasize certain semantic features of a scene because they seem critical with regard to the context in which they occur. The limitations of the flexibility within this system do, however, remain to be defined. In conclusion, we proposed here an important new hypothesis that suggests that the communicational demands of a situation and the speaker's intent (which is, of course, linked to these demands) regulate the operation of iconic gesture and speech in talk. Future research must provide the details of this operation.

Acknowledgements

Appendix

Transcription conventions:

Segments of speech analyzed are marked using 'single quotes'. The part of the verbal utterance that was directly accompanied by the iconic gesture is marked using [square brackets], thus the brackets indicate the start and end points of the iconic gesture with respect to the speech. The iconic gesture that accompanied the verbal utterance is described below the extract of speech in each case , and this description is also contained within square brackets. In the present study, we considered the iconic gesture as a whole in our transcriptions, whereas in the gesture transcriptions extracted from McNeill (1992) only the start and end point of the stroke phase of the iconic gesture is marked (see McNeill 1992: 12).

References

Allen, Robert E. (ed.) (1990). *The Concise Oxford Dictionary of Current English.* Oxford: Clarendon Press.

Beattie, Geoffrey and Aboudan, Rima (1994). Gestures, pauses and speech: An experimental investigation of the effects of changing social context on their precise temporal relationships. *Semiotica* 99 (3/4), 239–272.

Beattie, Geoffrey and Shovelton, Heather (1999a). Do iconic hand gestures really contribute anything to the semantic information conveyed by speech? An experimental investigation. *Semiotica* 123, 1–30.

— (1999b). Mapping the range of information contained in the iconic hand gestures that accompany spontaneous speech. *Journal of Language and Social Psychology* 18, 438–462.

— (2002). An experimental investigation of some properties of individual iconic gestures that mediate their communicative power. *British Journal of Psychology* 93, 179–192.

British Psychological Society (2000). *Code of Conduct, Ethical Principles, and Guidelines.* Leicester: The British Psychological Society.

Butterworth, Brian (1975). Hesitation and semantic planning in speech. *Journal of Psycholinguistic Research* 4, 75–87.

Clark, Herbert H. and Clark, Eve V. (1977). *Psychology and Language.* New York: Harcourt Brace Jovanovich Inc.

Clark, Herbert H. (1996). *Using Language.* Cambridge: Cambridge University Press.

Efron, David (1941). *Gesture and Environment.* New York: King's Crown Press.

Ekman, Paul and Friesen, Wallace V. (1969). The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica* 1, 49–98.

Ellis, Andrew and Beattie, Geoffrey (1986). *The Psychology of Language and Communication.* London: Weidenfeld & Nicolson.

Holler, Judith and Beattie, Geoffrey (2002). A micro-analytic investigation of how iconic gestures and speech represent core semantic features in talk. *Semiotica* 142 (1/4), 31–69.

Kelleher, Pat (2000). *Tom and Jerry* 30, 22–35.

McNeill, David (1985). So you think gestures are nonverbal? *Psychological Review* 92 (3), 350–371.

—(1992). *Hand and Mind: What Gestures Reveal about Thought.* Chicago: University of Chicago Press.

Schank, Roger, C. (1972). Conceptual dependency: A theory of natural language understanding. *Cognitive Psychology* 3 (4), 552–631.

Schank, Roger C. and Abelson, Robert P. (1977). *Scripts, Plans, Goals, and Understanding: An Inquiry into Human Knowledge Structures.* Hillsdale: Lawrence Erlbaum Associates.

Singer, Murray (1994). Discourse Inference Processes. In *Handbook of Psycholinguistics*, Morton A. Gernsbacher (ed.), 479–515. London: Academic Press.

Tower, Peter; Bryant, Bridget; and Argyle, Michael (1978). *Social Skills and Mental Health.* London: Methuen.

Judith Holler is a Ph.D. student at the University of Manchester <judith.holler@stud.man.ac.uk>. Her research interests include the interaction of iconic gesture and speech in conversational interaction.

Geoffrey Beattie is Head of the Psychology Department at the University of Manchester <beattie@psy.man.ac.uk>. His research interests include discourse analysis, nonverbal communication and its relationship to speech, and ethnography. His major publications include 'An experimental investigation of the role of iconic gestures in lexical access using the tip-of-the-tongue phenomenon' (with J. Coughlan, 1999), 'Mapping the range of information contained in the iconic hand gestures that accompany spontaneous speech' (with H. Shovelton, 1999), 'Do iconic hand gestures really contribute anything to the semantic information conveyed by speech? An experimental investigation' (with H. Shovelton, 1999), and 'Iconic hand gestures and the predictability of words in context in spontaneous speech' (with H. Shovelton, 2000).