Discourse Coherence and Gesture Interpretation

Alex Lascarides and Matthew Stone
School of Informatics, University of Edinburgh, alex@inf.ed.ac.uk
Department of Computer Science, Rutgers University, matthew.stone@rutgers.edu

MARCH 2, 2009

Abstract

In face-to-face conversation, communicators orchestrate multimodal contributions that meaningfully combine the linguistic resources of spoken language and the visuo-spatial affordances of gesture. In this paper, we characterise this meaningful combination in terms of the COHERENCE of gesture and speech. Descriptive analyses illustrate the diverse ways gesture interpretation can supplement and extend the interpretation of prior gestures and accompanying speech. We draw certain parallels with the inventory of COHERENCE RELATIONS found in discourse between successive sentences. In both domains, we suggest, interlocutors make sense of multiple communicative actions in combination by using these coherence relations to link the actions' interpretations into an intelligible whole. Descriptive analyses also emphasise the improvisation of gesture; the abstraction and generality of meaning in gesture allows communicators to interpret gestures in open-ended ways in new utterances and contexts. We draw certain parallels with interlocutors' reasoning about underspecified linguistic meanings in discourse. In both domains, we suggest, coherence relations facilitate meaning-making by RESOLVING the meaning of each communicative act through constrained inference over information made salient in the prior discourse. Our approach to gesture interpretation lays the groundwork for formal and computational models that go beyond previous approaches based on compositional syntax and semantics, in better accounting for the flexibility and the constraints found in the interpretation of speech and gesture in conversation. At the same time, it shows that gesture provides an important source of evidence to sharpen the general theory of coherence in communication.

Keywords: coverbal gesture, semantics, discourse coherence.

1 Introduction

People use their whole bodies in their joint effort to share their ideas with one another. They intend their actions to be understood as coordinated ensembles. So they adapt what they do with their hands and with their voice to ensure the synchronous performance of gesture and speech (see e.g., (Kendon, 2004, Ch 7) or (McNeill, 2005, Ch 2)). They repeat these embodied utterances when necessary, again with coordinated delivery of gesture and speech, to make sure their audience has attended (see e.g., (Kendon, 2004, Ch 8)). And when they must repair an utterance, they adapt both what they say and what they do in symmetrical ways (again see e.g., (Kendon, 2004, Ch 8)). Addressees, in turn, track the embodied utterances of their interlocutors as coordinated ensembles. For example, they understand and then use the visual or spatial information that appears only in their partners' gestures in their own subsequent contributions to the conversation (see e.g., (Cassell, McNeill, & McCullough, 1999)).

These integrative connections between speech and gesture give evidence of a deep underlying relationship. It suggests that interlocutors use a single fundamental set of interpretive principles to produce and understand ensembles of multimodal communicative action. In this paper, we propose to derive these interpretive principles from a general theory of COHERENCE in communicative action. Coherence theory is an approach to describing communicative action in terms of the interlocutors' BOUNDED RATIONALITY in coordinating contributions to conversation. The driving intuition is that interpretation must make sense of what a speaker is doing, by explaining why and how each constituent communicative action is SEMANTICALLY RELATED with the other communicative actions in the discourse. A COHERENT INTERPRETATION shows the speaker acting sensibly, presenting information that fits together semantically into an intelligible extended description, and doing so through an overall organisation and through choices of specific elements that make these semantic relationships clear. The ideal of coherence is a concerted programme of communicative action that gradually presents a complex idea to a group of interlocutors. However, coherence theory acknowledges that speakers may fail to meet this ideal. The strategies that actual human speakers exhibit, in negotiating, adapting and repairing their contributions as they struggle to express themselves, also fit coherence theory. Even in these cases, speakers draw on established patterns of presentation and interpretation to make clear what they mean and how it relates to what has been said before—in the case of communicative actions that repair prior contributions, coherence theory helps to identify the part of the prior actions whose interpretation should be discarded and the part whose interpretation should form a part of the current illocutionary act (P. Heeman & Hirst, 1995; Ginzburg, Fernandez, & Schlangen, 2007).

The theory of coherence originates in models of discourse interpretation in artificial intelligence and computational linguistics (Grosz & Sidner, 1986; J. Hobbs, Stickel, Appelt, & Martin, 1993; Kehler, 2002; Asher & Lascarides, 2003; Webber, Knott, Stone, & Joshi, 2003). Previous coherence models have treated language in isolation from other modalities of communication. For such models, the key question is how pragmatic principles interact with conventional, syntactically-articulated, symbolic meanings, as determined by a linguistic grammar. To account for embodied communication, however, we must revisit these pragmatic principles in light of the improvised, holistic and iconic signification that are characteristic of gesture.

We suggest in this paper that this interchange leads to new insights both into the interpretation of gesture and into the pragmatics of coherent communication. There are deep parallels between descriptive accounts of the contribution of gesture to embodied conversation and models of the contribution of spoken utterances to coherent discourse. Thus, by linking gesture interpretation to the theory of coherence, we get a new perspective that helps explain why we should see gestures interpreted with the flexibility we do, and with the invariants and constraints we do. We can even build on the tradition of coherence theories to describe gesture interpretation in formal and computational models. At the same time, by tying the theory of coherence specifically to gesture interpretation, we also gain new sources of evidence to sharpen the theory. Analysis of gesture demands that theories account for the full range of possibilities for meaning-making, including the many devices and strategies whose central place in communication is obvious in embodied conversation but easily underestimated in studies of purely linguistic discourse.

In our presentation, we specifically explore TWO key principles of coherence theory as they apply to the interpretation of gesture combined with speech. The first principle is that communicative actions are organised into hierarchical structures of semantically-related units. The semantic relationships between units are called COHERENCE RELATIONS, and include such relationships as ELABORATION, CAUSE–EFFECT, CONTRAST, and even REDUNDANCY (that is, repeated presentation of the same information) and CORRECTION (that is, presentation of new information that is intended to substitute parts of a prior contribution). Coherence relations structure communicative actions by showing how the speaker is grouping ideas together to highlight the meaningful relationships among them. These relationships give evidence of the speaker's intentions in the discourse; they track the relationship among ideas as interlocutors build up a line of argument, revise it, or even abandon it. They therefore serve as a bridge between semantics and pragmatics.

In Section 2, we introduce coherence relations, and explore the usefulness of using models of coherence relations in theories about the contribution of gesture in the interpretation of embodied utterances. Using established coherence relations from discourse, including relationships such as ELABORATION and CAUSE–EFFECT, provides an attractive framework to make precise the familiar principle that speech and gesture "combine to express a single thought"—in other words, that the content conveyed by gesture fits with the content conveyed in speech as an integrated part of the speaker's overall message (McNeill, 1992; Bavelas & Chovil, 2000; Engle, 2000; Kendon, 2004). However, this application also requires us to be more precise about coherence relations, since (for example) gesture highlights the importance of relationships of DEPICTION and REDUNDANCY.

The second key principle of coherence is that inferring coherence relations among communicative actions can help interlocutors to augment the meaning of a speaker's act as revealed by just its form with the content that the speaker intended to convey. In particular, such inferences specify ways that a unit of discourse can continue to contribute information about the same entities described in structurally-related units, and use the same communicative patterns to do so; in this way, coherence models the influence that salient information in the context has on interpreting utterances in extended discourse. This yields a more specific, contextually relevant, interpretation of the communicative act than is derivable from just its form.

In Section 3, we explore how the inferences that interlocutors use to establish coherence specialises the abstract possibilities for gesture meaning to its particular discourse context. One source of this coherence is the relationship between gesture and simultaneous speech. We will argue that the inference that resolves underspecified meaning by relating communicative actions across modalities, as described for example by Kendon (2004), parallels the inference that resolves underspecified linguistic meanings in successive utterances. Another source of coherence is the relationship between successive gestures. Successive gestures often continue with established figurations in representing objects, space, events and actions (Emmorey, Tversky, & Taylor, 2000; Haviland, 2000; McNeill et al., 2001). This strategy broadly parallels the consistent resolution of underspecified linguistic meanings, conditioned on discourse structure and coherence (Grosz & Sidner, 1986; Webber, 1991; Asher, 1993; Brennan & Clark, 1996; Webber et al., 2003). Assimilating these interpretive links to other cases of establishing coherence clarifies the fundamental principles involved. In particular, we show how the processes for identifying coherence relations, resolving semantic underspecification, and determining interpretive connections in embodied utterances can exploit possibilities that have not previously been recognised in coherence theories.

We deepen our approach and explore its consequences in Section 4 by discussing an extended example from a naturally-occurring conversation. This discussion illustrates the fundamental contribution of our work: to serve as a bridge between descriptive approaches to gesture and formal and computational approaches to conversation. We emphasise how a close descriptive reading of the conversation can be informed by an analysis that exploits the theory of coherence, and how the example in turn motivates new distinctions and questions for the theory.

Previous computational models of gesture interpretation (Johnston et al., 1997; Johnston, 1998; Cassell, 2001; Kopp, Tepper, & Cassell, 2004) have not appealed to coherence as a general principle. Instead, they have taken specific semantic relationships to be constitutive of the natural combined use of speech and gesture. That fits many cases, such as the demonstrative gestures that cospecify referents for deictic expressions in language. But it does not fit all; many utterances show speech and gesture in a looser, inferential relationship. Without coherence, prior accounts are both too restrictive, because they fail to acknowledge the diverse ways speakers can communicate coherently, and too general, because they fail to predict the specific resolutions of content in context that can be required to establish coherence. The present work therefore promises to pave the way for future implementations that better realise the expressive richness of our own communicative action. We contrast our approach more precisely with previous formal models in Section 5.

Fundamentally, our agenda is to demonstrate that principles of pragmatic interpretation are general, applying to the interpretation of communication in whatever medium it takes place. On this view, what makes language special is the specific devices it offers for meaning-making (i.e., the precise predicate argument structure that's borne from lexical subcategorisation and syntactic dependencies); not the meanings themselves or the ways in which those meanings resolve to a pragmatically preferred interpretation in context. Likewise, what makes gesture special is its specific devices for meaning-making, such as its distinctive iconic representation and the real, virtual and metaphorical spaces which speakers can evoke through gesture. We close in Section 6 by outlining the challenges that remain in reconciling the methodologies and insights explored in different approaches to pragmatic interpretation—charting a course towards the characterisation of a common pragmatic substrate for interpreting collaborative communicative action that figures in language, in gesture, and in their use in combination.

2 Coherence Relations

Coherence relations offer an inventory of things that a speaker might be doing by performing communicative actions in conversation. The term "coherence relation" serves as a reminder that speakers typically do not present information in isolation; rather, they expand on what they have already contributed, by elaborating, explaining, continuing a narrative, drawing a contrast, and so forth. Interpreters expect speakers to organise discourse to highlight these meaningful relationships among successive contributions, and examples like (1), as discussed by Hobbs (1979) and Kehler (2002), show how far interpreters go to draw inferential connections between juxtaposed material as part of establishing discourse coherence.

(1) a. John took a train from Paris to Istanbul. He has family there.

b. John took a train from Paris to Istanbul. He likes spinach.

Discourse (1a) makes sense. Visiting family gives a natural reason for John to make the trip, and it's natural for a speaker to continue talk of John's trip by giving an explanation for it. On the other hand, the juxtaposition of the two sentences in (1b) is mysterious, and so the example is unsatisfying as a discourse. Even though both sentences offer straightforward descriptions of John, interpreting (1b) leaves one with a feeling that something is missing—perhaps some exceptional situation that would make John go to Istanbul for spinach. What's missing, following Hobbs (1979), is COHERENCE:

> ...the very fact that one is driven to such explanations indicates that some desire for coherence is operating, which is deeper than the notion of a discourse just being "about" some set of entities. (J. R. Hobbs, 1979, p 67)

A particularly powerful indicator of the interpretive effects of coherence is the resolution of reference. For example, in (1a), both Paris and Istanbul are mentioned in the first sentence, but we take the anaphoric expression *there* in the second sentence to refer to Istanbul. If we make these alternative readings explicit, by replacing the second sentence with *He has family in Istanbul* or *He has family in Paris*, we continue to find the first more natural. By the same token, we ordinarily expect an explanation of a trip to account for its destination. So the resolution of *there* to Istanbul MAXIMISES COHERENCE: it allows us to understand the speaker of (1a) as acting in the most orderly and intelligible possible way.

Referential interpretation thus provides a diagnostic for the different relationships that can make discourse coherent. Consider (2), taken from Kehler (2002) in a discussion that builds on Lascarides and Asher ((1993)) and Webber (1988).

(2) a. Max spilt a bucket of water. He tripped on his shoelace.

b. Max spilt a bucket of water. He spilt it all over the rug.

c. Max spilt a bucket of water. John dropped a jar of cookies.

In (2) the TEMPORAL REFERENCE associated with the past tense verb in the second sentence varies across the examples. Max's tripping, as described in (2a), PRECEDES the spill. The inundation of the rug in (2b), however, describes that original spill in a larger context, so the two sentences report SIMULTANEOUS happenings. Finally, (2c) seems neutral about temporal order among the two events it describes.

This variation is just what's required to support the natural underlying inferential connections between the utterances in these different discourses. In (2a), the trip EXPLAINS the spill. Explanation is an instance of a class of coherence relations Kehler calls CAUSE–EFFECT, and of course a cause must precede its effects. The elaboration of (2b) is a case of coherence relations of CONTIGUITY: juxtaposing information because it takes place at a common space or time. Finally, (2c) reports a parallel between Max and John, and exhibits a general tendency to organise discourse so that eventualities with a shared RESEMBLANCE are described together. In this case, the parallel suggests that a speaker of (2c) would aim to draw a general conclusion that Max and John were both klutzy during some independently established reference time—an inference which need not trigger any further inference about the relative times when the spill and the drop actually happened.

Referential interpretation becomes a particularly powerful diagnostic of coherence when we broaden our enquiry from written texts like those in (1) and (2) to the improvised and interactive contributions speakers make to spontaneous dialogue. Consider Strawson's famous interchange in (3), for example (1952, p 187):

(3) a. X: A man jumped off a bridge.

b. Y: He didn't jump, he was pushed.

Y's utterance in (3b) functions as a CORRECTION of *X*'s utterance in (3a). We can identify this coherence relation in part by the fact that we take *He* in (3b) to evoke the individual that *X* intended to describe as jumping in (3a). Recognising *Y*'s contribution is a denial or correction of *X*'s also depends on the spatio-temporal reference of the tenses in their utterances being made equal. This resolution is exactly what's required to understand *Y* as disagreeing with *X* and as offering an alternative explanation of the matter at issue. Speakers often provide content that denies earlier content in the conversation, so the possibility of examples such as (3) is not surprising. Indeed, despite their problematic nature, such dialogues continue to showcase speakers acting to present new content in a recognisable relation to its context, and so coherence theory can describe these dialogues as readily as it describes cases where interlocutors accept and build on one anothers' contributions. (Lascarides & Asher, 2009) offer a more extensive discussion of correction in coherence theory, including a formal treatment. Their treatment demonstrates how the semantic representation of a dialogue that features a dispute remains consistent, even when a speaker denies the content of his own prior assertions. Such consistency in the analysis of the conversation is required in order to maintain consistent predictions about what's agreed upon when a dispute has taken place—observe in (3) that *X* and *Y* both agree that the man went off the bridge, and the dispute centres on how this happened.

Similarly, consider (4) from (P. A. Heeman & Allen, 1999, Example 36, p 568):

(4) the engine can take as many um it can take up to three loaded boxcars

This is a disfluent utterance, produced by a speaker in an experiment on collaborative problem solving. Two partners were tasked with scheduling a set of deliveries in a train transport network. The utterance in (4) describes one of the constraints the pair encountered in developing their plan: a limit on how much an engine can carry. The utterance involves a false start. The speaker initially intends to express the limit with a noun phrase beginning *as many*, but abandons this utterance and instead formulates another with *up to*. The example invites us to see REPAIR as a coherence relation. When we understand what the speaker is doing here, we understand that the speaker intends to adapt a provisional utterance *the engine can take as many* in favour of an alternative realisation *it can take up to three loaded boxcars*. Crucially, here, *it* in this example is intended to corefer with *the engine*. Had it not been for the constituent under repair, this interpretation for *it* would likely have been unavailable in context. Here, it's just what's required to understand the second half of the utterance as providing a definitive restatement of the idea provisionally and partially broached in the first half. As in (3), then, the example shows the programmatic, uniform account coherence theory gives to common patterns of utterance use in conversation, even problematic ones. In fact, existing treatments of repair in coherence theory remain largely exploratory, though promising work has been done for the special cases of the negotiation of meaning through interaction (P. Heeman & Hirst, 1995; Ginzburg et al., 2007), and specifically on the interpretation of utterance fragments (Schlangen, 2003; Lücking, Rieser, & Staudacher, 2006).

We suggest that coherence relations continue to help describe the inferential character of interpretation when we broaden our enquiry further still and consider gesture. Specifically, we argue that coherence relations can play a similar role in the analysis of gesture interpretation as in the analysis of discourses such as (1–4). We can use coherence relations fruitfully both to help characterise what speakers do when they contribute content to conversation using gesture, and to help motivate the interpretive inferences that are required to make sense of gesture.

It has long been recognised that speakers can orchestrate embodied utterances in which gesture and speech present complementary information that fits together into an integrated whole. Kendon's discussion of (5) epitomises this type of descriptive analysis (2004, Ex 7.1, p 114).

```
(5)                                             0.3 sec
        M: he used go down there and throw (..........) GROUnd rice over it
        RH                   | prep    | stroke        | recovery   |
                             | GESTURE PHRASE          |
                             | GESTURE UNIT                         |
```

This utterance is an extract from speaker M's account of his father's methods for ripening cheeses for sale in his grocery; *he* is M's father and *it* is a representative cheese. The utterance is delivered with a gesture of the right hand. In the preparatory phase of the gesture, the speaker positions his right arm in gesture space, with the right forearm forward, the palm upward, and the hand in a loose fist. In time with the word *throw*, and the pause following it, the speaker shakes the hand outward twice from the wrist. In context, the gesture appears to exemplify the action his father would take in scattering a handful of powder over the top of the cheese. In this interpretation of the speaker, we take both the demonstration and the verbal description to characterise the speaker's father's action, and we arrive at an understanding of the type of action performed that respects both the verbal account as *throwing ground rice over the cheese* and the gestural depiction as an action of scattering powder.

We believe examples such as (5) yield to an insightful account in terms of coherence relations. From this perspective, establishing the interpretive connection between gesture and speech in (5) involves recognising that the speaker is using the gesture to depict additional aspects of the situation described in the simultaneous speech. This is a kind of CONTIGUITY relationship. Recognising this coherence relationship shapes the interpretive inferences through which we fit together our understanding of M's gesture and speech. For example, we see the gesture in this discourse context as a depiction of literal action and not as a metaphor. (In some cases, gestures apparently similar to the speaker's here, in which the hand is shaped to hold or convey a quantity of stuff, are interpreted as metaphorical depictions of ideas grasped or communicated through language (McNeill, 1992, p 147).) Conversely, in light of the speaker's depiction, we understand the action of *throwing* in a very general way, simply as *casting through the air*, without its stereotypical implications as indicating forceful propulsion of material, by jerking the arm straight, over the shoulder.

Nevertheless, examples such as (5) do not on their own provide a good argument for coherence theory. Coherence theory offers an inferential account where the interpretation of one communicative act may relate only indirectly to that of another, and it fundamentally depends on the speaker's overall purposes in the discourse. Cases like (5) seem compatible with a much more constrained approach to the joint interpretation of gesture and speech. In (5), the gesture signifies through a transparently iconic portrayal of an action, and transparently depicts the topic of the associated speech. Moreover, as formulated for discourse, coherence theory offers few insights into the kind of presentation that this gesture represents. Illustrating something described in associated words is a kind of communicative action, and its relational form fits the scheme of coherence theory. But the action of illustration is new to coherence theory. It seems closely bound up with the iconic signification of the gesture. And superficially similar relationships that have been described for discourse, such as PARAPHRASE or RESTATEMENT, are actions that speakers would normally use with different intentions and different functions, under very different circumstances to a gesture like (5). A restatement of an idea recently presented in discourse, for example, is regularly presented in RHETORICAL OPPOSITION to countervailing ideas that have been presented in the interim (Horn, 1991; Walker, 1993).

Similarly, in our view, the distinctive use of gesture in disfluent utterances speaks neither for nor against a coherence theory of the relation of gesture and speech. Many researchers have been struck by the apparent role of gesture in facilitating problematic lexical access, as in (6) below (see (Krauss, Chen, & Chawla, 1996) for a review). The utterance in (6) occurs shortly after (5), and introduces a story which now describes how the speaker's father would test if a cheese was ready to sell, by boring out a sample (Kendon, 2004, Fig 9.5A, p 171):

(6) **M**: An' he got like ehm an auder an auger
M's right hand models the instrument while his left hand models the cheese..

By the time the speaker says *ehm*, he has already adopted the pose of the gesture, with just the index finger of the right hand extended, pointing vertically downward, and the left hand holding an open fist.

We can give an analysis of this utterance that parallels the analysis of the disfluency in (4). In (6), the speaker incrementally constructs a complex utterance, where part of the verbal material is abandoned (the phrase *an auder*), another part gives a replacement (the phrase *an auger*), and meanwhile an associated action, performed across the utterance as a whole, depicts an aspect of a related situation (the typical use of an auger). This analysis explains what the speaker is doing as he presents these bits of content in relationship to one another. It shows how coherence theory can give an interpretation to disfluent examples, even though they may be ungrammatical and, in fact, involve nonwords. For many researchers, however, what's interesting about such examples is not the interpretation we finally arrive at. It's the possibility that the speaker's expressive embodied action may actually assist him in retrieving the word *auger*—or that it may assist the addressee to recognise what the speaker means even before the correct word is retrieved. Without further assumptions about the processes of production or understanding—issues whose relevance is emphasised by (Krauss, Chen, & Gottesman, 2000)—coherence theory neither predicts nor precludes such effects.

Our argument for coherence theory, then, centres on cases where gestures are not transparently related to the words they accompany, and where the inferences that connect the two resemble connections often found in discourse. We offer three illustrations here, drawn from (Engle, 2000; Kendon, 2004; McNeill, 2005). The attested examples we describe here have been chosen to emphasise the systematic descriptive work which leads analysts to interpret gestures as carrying specific information which is not a portrayal or straightforward elaboration of what is described in simultaneous speech. The investigators differ in how they characterise the looser relationship they find between gesture and speech. For Engle (2000), these examples show that gestures and speech are not to be interpreted as separate channels presenting corresponding information, but as composite signals in semantic interaction. For Kendon (2004), they show that gestures can convey implicit, inferable or broader concepts not evoked explicitly in speech. For McNeill (2005), they show that the conceptualisation underlying a multimodal communicative action may itself be complex, with partial expression in speech and partial expression in gesture. We see coherence as a new, more finely-dileneated characterisation of these relationships: coherence relations provide a theoretical framework to describe how related communicative actions supplement one another, providing a complex description whose components bear inferential relationships to one another, interact semantically, and gel into an integrated argument. At the same time, the analysis represents a challenge to coherence theory, because it calls for a more systematic analyses of relationships such as depiction and illustration that connect gesture and speech in examples such as (5).

Consider utterance (7), drawn from Kendon's fieldwork in Naples (2004, Ex 29, p 181). Kendon offers the example to illustrate how gesture sometimes "serves to specify a dimension of reference that is not itself directly given in words" (2004, p 181).

(7) *Giova' m'eva i' a fa nu parë 'i scarpë (.) io aggia vistë*
purë nun è cosë pëcchë troppë carë.
Giovanni I was going to go buy a pair of shoes I had seen,
but it wasn't the case because too expensive.
As [Peppe, the speaker] says "troppë carë" ('too expensive') he lowers his open right hand, palm vertical, twice toward his left hand, held open, palm up.

Peppe's gesture here is an instance of *na mazzata*, a distinctive Neapolitan gesture with a largely conventionalised interpretation—a "narrow gloss" gesture about whose interpretation speakers have very precise intuitions. The gesture is a metaphorical demonstration of a blow from a bat or club, and dramatises how one feels after encountering something unexpectedly unpleasant—one feels as though one has been hit. In using *na mazzata* here, Kendon writes, "Peppe shows that the discovery of the high price of the shoes was a shock for him" (2004, p 183). On this understanding, the gesture is NOT a visualisation of the shoes or their price—the individuals explicitly evoked in the accompanying speech. Instead, the gesture characterises something related: the speaker's shock at discovering the state of affairs conveyed in the speech. This reaction and the state of affairs that evokes it stand in a CAUSE–EFFECT relationship. The same sort of relationship often goes unstated between successive sentences in narrative discourse, where, for example, the sentence "I was shocked" would normally be understood to describe the speaker's reaction to an event described just prior.

Engle (2000, Table 8, p 37), meanwhile, used a mismatch in the number of objects portrayed in speech vs. gesture to characterise (8) as a case where speech and gesture must be interpreted in inferential relation to each other.

(8) They [have **springs**.]
Speaker places right pinched hand (that seems to be holding a small vertical object) just above left pinched hand (that seems to be holding another small vertical thing).

Engle asked her subjects to explain the workings of an ordinary cylinder lock, starting from Macauley's visual explanation from *The Way Things Work* (1988, pp 16–17). The diagrams label the cotter pins on the lock, depict how they extend down into the cylinder and hold it in place while the door is locked, and show how they are raised out by the profile of the key. As the speaker suggests by the ensemble of communicative actions in (8), where *they* refers to the cotter pins, each pin is held down not only by gravity but by the action of a spring coiled around it and anchored to the top of the lock mechanism. Note then that the gesture is not directly a visualisation of the spatial relationship between the SET of cotter pins and the SET of springs, as they are evoked in the words of (8). Instead, the gesture shows the vertical arrangement of ONE representative pin and its corresponding spring. This EXEMPLIFICATION relationship is also frequently seen in discourse, when a speaker presents first the statement of a generalisation and then an account of a particular instance or instances. Indeed, identifying that this coherence relationship holds and disambiguating the plural predication in the linguistic phrase to a DISTRIBUTIVE INTERPRETATION are logically co-dependent. The ensemble in (8) is also a case of RESEMBLANCE, the parallel description of related situations to exhibit similarities and differences among them. At the same time, of course, (8) also involves the relationship of depiction which is distinctive to gesture.

Finally, McNeill (2005, Ex 4.3.1–4.3.5, pp 139ff) offers an account of the indirect relationship between gesture and speech in (9).

(9) a. [top bi sekil-de] ball in one way
hands hopping in place

b. [zipla-ya zipla-ya] while hopping
hands hopping and moving right

c. [yuvar-lan-a yuvar-lan-a] while rolling itself
hand moves right

d. [sokak-tan] on the street
hands again move right without hopping = path alone

e. [gid-iyo] goes
hands again move right without hopping = path alone

This is a Turkish speaker, narrating an episode from a Tweety and Sylvester cartoon. Sylvester attempts to reach Tweety by climbing up a drainpipe. Tweety foils the plan by dropping a bowling ball down the drainpipe. The bowling ball winds up inside Sylvester, and, as the speaker describes in (9), the ball—working impossibly, from the inside!—rolls Sylvester down the street. McNeill offers this idiomatic translation of the entire utterance: the "ball somehow, hopping, rolling, goes on the street".

The example in (9) was first described by Özyürek (2001), who argues convincingly that the gestures in (9a), (9b), and (9c) do not depict the same actions or events that the accompanying speech describes. Özyürek's conclusion capitalises on the methodology developed by the McNeill lab. Since speakers base their narration on a stimulus presentation, analysts have independent access to the source for the speaker's gestures. When we interpret the gestures of (9) in relation to the original cartoon stimulus, setting aside the accompanying speech, we find that the gestures glossed as *hopping* seem to match the flailing animation of Sylvester's legs in the air as the bowling ball rolls him forward from underneath, while the rightward movement matches the overall dynamic of cat and ball, as effected by the ball's inertia.

For McNeill, this extended utterance is a coherent case of co-expressive speech and gesture because it represents the speaker's attempt to portray the CAUSE–EFFECT relationships behind Sylvester's motion down the street. This might be seen as an analogue to the explanatory CAUSE–EFFECT relationship that implicitly connects the two sentences of (2a). In (9), the first phrase describes the ball acting somehow, and is accompanied by a gesture that depicts the effects of this action on Sylvester. The second phrase indicates the hopping motion while depicting the broader causal interaction between the hopping cat and the ball driving him forward. The third phrase shows describes the rolling and depicts the motion down the street that is its causal result. McNeill concludes that "[t]he entire sentence, as it unfolds, embodies the speaker's analysis of the causal structure concerning the way the cat ended up with a bowling ball inside him and rolling down the street" (McNeill, 2005, p 141).

The gestures in (7–9) provide particularly clear evidence for the coherence of speech and gesture because their experimental and analytical contexts give such strong evidence for the specific interpretation for gesture and the specific interpretation for speech. In most cases, we suspect, we can account for the semantic relationship between gesture and speech so precisely only in the context of a deeper analysis of the forms and meanings available in each modality to represent the world. In the next section, we explore these form–meaning mappings in more detail. We find not only further parallels between gesture and speech but also further illustration of the perspicuous accounts of the integrated interpretation of speech and gesture that can be articulated within the framework of coherence theory.

3 Interpretive Inference

Coherence theory describes utterance interpretation in terms of the specific purposes that underlie specific communicative actions. It emphasises that these intentions often involve substantially more detail and precision than is available in general from utterance meaning as revealed by just its form. For example, consider the utterance (10), as described in (Stone, 2004a, Ex (4a), p 42):

(10) So are we all set?

In the abstract, *so* invites the addressee to consider some issue in light of an interaction just completed; *we* can denote any group containing the speaker; and *all set* describes the conclusion of any process of preparation. This meaning is fundamentally underspecified and the underspecification disappears when the utterance is used in a particular context—say, in a setting where two interlocutors are cooking dinner together. Either can then use (10) to ask the other whether the two of them have, as a result of the activities that they have just accomplished, completed their work in readying the meal. We say that this interpretation RESOLVES the abstract meaning to specific values that give the utterance its relevance to the context. This resolution supplies the two collaborators as *we*, the dinner as the objective for which they may be *all set*, their ongoing work and interaction as the circumstances for which *so* prompts reconsideration.

Coherence theory describes this resolution as the satisfaction of constraints provided by utterance meaning on the one hand and semantic links to the discourse context on the other. Each constraint instructs the addressee to determine the relevant instantiation of a more abstract meaning by finding an appropriate specific value. Candidate values should be grounded in the model of DISCOURSE CONTEXT, a record of the information that speakers have explicitly contributed thus far (Lewis, 1979; Thomason, 1990; Poesio & Traum, 1997). And they should respect DISCOURSE STRUCTURE, a hierarchical organisation of the interaction into successively larger spans or SEGMENTS that coherently address related issues (Grosz & Sidner, 1986; Mann & Thompson, 1987). However, appropriate values may also be calculated by limited forms of inference from commonsense background knowledge.

Resolutions are determined holistically, because multiple constraints may govern the same values and all must be satisfied simultaneously. In addition, the overall resolution must lead to a coherent interpretation, which gives a satisfactory explanation of the speaker's action in presenting this information now. Thus, in coherence theory, resolving interpretation and establishing coherence are not two separate processes that might be undertaken in stages but two complementary perspectives on a single process of utterance understanding. The perspective of resolution highlights INTERPRETIVE PREFERENCES that distinguish resolutions of meaning that interlocutors prefer to rely on because they are clear and easy to understand. These preferences complement coherence relations in bridging semantics and pragmatics; they can guide speakers' choices in planning coherent utterances and listeners' inferences in reconstructing specific interpretations in context.

In general, these interpretive preferences favour the reuse of salient information that has been presented in related units of discourse. One way to reuse information is simply to recover entities that have been explicitly evoked in the discourse context, as suggested in the resolution of (10). Another is to follow an interpretive precedent established earlier in discourse, to resolve related meanings in parallel ways (Lewis, 1979; Brennan & Clark, 1996). This is illustrated in (11), abridged from an obituary of sculptor Oscar Lenz (Levy, 1913, p 78):

(11) Lenz came to New York to study under Saint Gaudens, then went to Paris where he studied under Saulierre.

In (11), the use of *came* in the first clause establishes a precedent that the deictic locus for the discourse is New York. This precedent remains in effect thereafter, so that we understand the subsequent use of *went*, to describe Lenz's move to Paris, with reference to the same deictic locus. For a speaker in Paris, the first event might be *going*, the second *coming*.

A final way to reuse information is to base an inference on information contributed in prior discourse, in conjunction with general commonsense background knowledge (J. Hobbs et al., 1993). This is illustrated in (12).

(12) a. Slide sleeve onto elbow.

b. Reposition sleeve.

Text (12) is taken from a repair manual for the fuel system in a military aircraft (USAF, 1988), as modeled computationally by (Stone, Doran, Webber, Bleam, & Palmer, 2003). These instructions describe part of a repair procedure where vents in the aircraft are joined together. Normally, a sleeve is used to seal the connection and a coupling is used to hold the vents and sleeve in place. To gain access at the beginning of a repair, personnel slide the sleeve out of the way, often as in (12a) as far as the next curve in the ductwork (called an elbow). Then, when the repair is done, they slide the sleeve back so that it again seals the connection between the vents, as in (12b). Thus, the action described in (12b)—despite its different vocabulary—is just the reverse of the action described in (12a). The meaning of (12b) in isolation of its context is, of course, underspecified. It describes any action which brings the sleeve to a definite location where it has been before. The intended resolution in this context, however, matches interpretive preferences: via inference it is linked to both the function of the sleeve (to seal the connection) and the fact, saliently evoked in the context of the instructions, that the sleeve was rigged to do this at the beginning of the repair. The precision of this information in guiding the resolution of the meaning of (12b) seems to explain why the manual uses this description in lieu of other possibilities (Stone et al., 2003).

Across all these different ways to reuse information, we find a further preference to exploit discourse structure in focusing attention on the information potentially most relevant for interpretation. The theory of coherence relations characterises each utterance in discourse as continuing the treatment of an issue provisionally developed across a unit of the prior discourse. In this case, the new utterance ATTACHES there in discourse structure and joins with this unit to form a larger discourse segment. In such circumstances, we prefer to resolve meanings against the discourse segment—at a level that engages with the overarching issue developed throughout the unit where the utterance attaches, rather than engaging with a narrower focus addressed only in part of this unit, or a broader focus that ties the current unit to still larger units of discourse. In (10), this preference fits the resolution of *so* and *all set* in the context of the ongoing activity. In general, this preference is at play in resolving the meanings of cue words such as *instead* or *for example* that speakers use to help specify how an utterance contributes to the discourse (Webber, 1991; Webber et al., 2003). But it also constrains the interpretation of a range of other references, including definite descriptions of entities and implicit reference to propositions and events (Grosz & Sidner, 1986; Asher, 1993).

We advocate describing gesture meaning, like linguistic meaning, through coherence theory's process of holistic resolution of constrained but underspecified meanings to specific values in context, guided by preferences to reuse salient information from related communicative actions. As with coherence relations, this proposal stays quite close to accounts of gesture meaning from the descriptive literature. For example, Kendon ((2004, p 169)) describes the gestural depiction of *throwing ground rice over the cheese* in (5) in terms of an abstract meaning that gets a specific interpretation in virtue of its relationship to simultaneous speech:

> Once again we may note that the action of this gesture phrase cannot be precisely interpreted until it is perceived as part of the gesture–speech ensemble in which it is employed. The action has a general, abstract significance which is made specific by the verbal component with which it is associated.

Nevertheless, as with coherence relations, this proposal calls new attention to the INFERENTIAL character of the resolution of gesture meaning. It emphasises on the one hand that we can characterise elements of gesture form as instructing the addressee to supply specific kinds of information in interpretation, and on the other hand that we can trace the specific sources for this information in the discourse context. Again, while this inferential account does seem to characterise what is involved in interpreting (5), such examples do not motivate the flexibility of the framework or the applicability of an analogy to linguistic meaning. Perhaps we need only a small inventory of ways gesture can represent, which can be selected in context based on simple constraints of co-expressiveness with associated verbal material. For example, in (5) the abstract form of the gesture can be resolved to a specific interpretation by recognising that the speaker has adopted a CHARACTER VIEWPOINT (McNeill, 1992, p 118ff) to give a schematic enactment of the activity simultaneously described in words. Moreover, since gesture typically involves iconicity or deixis, rather than the arbitrary signification characteristic of language, perhaps new mechanisms must be at work in pinpointing the specific interpretation of a gesture. We could postulate, for example, that a speaker's hand in a gesture has an abstract meaning that instructs interpreters to discover something that the hand represents. And when we abstract over the specific things a hand might depict, we might appeal to principles of natural correspondence that have no analogue in linguistic meaning.

To argue for extending the interpretive inferences of coherence theory to gesture, then, we draw on illustrative examples where gestures' abstract meanings and salient resolutions do exhibit a clearer parallel with the coherence of verbal material. We begin with (13), described as Example 93 and Fig 13.4 of (Kendon, 2004, pp 253–254). The utterance shows an abstract gesture whose meaning is very naturally described as an instruction to recover salient information of a specific kind, much as we saw in (10).

(13) *ci metto o dei pomodorini o un po' di passata, veramente*
I put either some little tomatoes or a little tomato purée, truly
As the speaker says "po' di passata, veramente", she lifts her left hand up to her shoulder, hand open, fingers vertical, palm facing her interlocutor.

In context, this gesture shows that the speaker stops herself from adding too much tomato purée, and emphasises that only a small quantity, and no more, is required for this recipe. The gesture of (13) illustrates what Kendon calls the *vertical palm, open hand prone* (VP) gesture. Deployed close to the speaker's own body, as here, it functions as a kind of stop sign marking a halt that the speaker puts to an activity. Thus this meaning encodes a underspecified constraint—which activity should be stopped? Thus the hearer must retrieve a specific activity of the speaker's and use it to interpret the gesture. In this case, the hearer needs to recover the potentially open-ended process by which the speaker incrementally measures out tomato purée to be added to her sauce. This resolution draws on commonsense inference and salient information in the related discourse. What we recover is the PREPARATORY PROCESS of the complete event the utterance specifically describes, where the speaker adds a specific quantity of the ingredient to the sauce. This causal part–whole inference is also common in the inferential interpretation of tense and aspect in natural language discourse (Moens & Steedman, 1988; Webber, 1988). The result is a meaning for the gesture that, in the terms of coherence theory, ELABORATES on the positive meaning of the verbal material, by suggesting that the speaker has a "self-imposed limit beyond which she does not go" in adding tomato to her sauce (Kendon, 2004, p 253)—and, in so doing, helps to give a more precise interpretation to the vagueness of her words *un po'* (a little) and *veramente* (truly).

Our next example highlights how a gesture, like a spoken sentence, can involve multiple dimensions of underspecified meaning that must be resolved jointly into a specific interpretation. In (14), from Fig 15.6 of (Kendon, 2004, p 322), the speaker of (5) is describing the special cake his father's shop would sell each year at Christmas time.

(14) a. and it was [pause 1.02 sec] this sort of [pause 0.4 sec] size
during the pauses, the speaker frames a large horizontal square using both hands; his index fingers are extended, but other fingers are drawn in, palms down.

b. and [he'd cut it off in bits]
the speaker lowers his right hand, held open, palm facing to his left, in one corner of the virtual square established in the previous gesture

We focus on the gesture in (14b), which as analysed by Kendon "shows just where and how the grocer would cut off a piece of the cake" (2004, p 323). In other words, like (8), it gets its coherence from depicting the spatial relationships involved in a REPRESENTATIVE EXAMPLE of a set of events explicitly referenced in the associated speech. Arguably, the gesture of (14b) involves at least THREE separate dimensions of form that each contribute their own abstract constraints towards interpretation. These constraints are resolved drawing on different information so as to yield an overall consistent coherent interpretation. First, there is the attitude and motion of the speaker's arm. There are an unlimited number of properties that an entity (which is not necessarily an arm) can have that can be depicted through this particular attitude and motion of the arm; and the properties so depicted can invoke a varying number of participants. But in this discourse context, the gesture is interpreted in a quite specific way: it mirrors the grocer's bodily action in cutting the cake, and thus is interpreted as an iconic depiction in character viewpoint. The hand, however, with its flat open shape, seems to depict not the grocer's hand, but the cutting implement that, we know by commonsense inference, the grocer must have used when he cut cake (for the instrument that's used to cut the cake isn't mentioned in the speech). Thus it is interpreted as MODELING an object, and this object is recovered from the discourse context by the general commonsense inference from a cutting to a blade. (Kendon (2004, pp 161ff) describes other cases where this speaker mixes character viewpoint action with the use of the hand to model an instrument.) Finally, the specific placement of the hand seems to index a specific cutting plane within the same virtual space in which the cake itself is deictically located in the gesture of (14a). This resolution might be seen as a case of repeated reference to a deictic frame from previous discourse, much as we saw with *came* and *went* in (11). Thus we must recognise deictic meaning in gesture as reusing salient information when it is interpreted with respect to a previously established virtual space (Emmorey et al., 2000; Haviland, 2000). The parallel with (11) is further strengthened if we assume that gestures may not only be related by coherence relations to associated speech, but may simultaneously stand in suitable coherence relations with one another, especially when, as in (14a) and (14b), they represent congruent extended depictions. Then we can take the interpretation of (14b), in reusing information from the related communicative action, like the interpretation of *went* in (11), as the preferred resolution of its deictic meaning.

We offer (15), described by McNeill in Examples 7.18ff of (1992, p 191) and Table 4.2.1 of (2005, p 118), as a further example of the role of discourse structure in guiding an integrated resolution of independent elements of gesture meaning to salient information in context.

(15) a. and as he's coming up
Accompanied by an observer viewpoint iconic gesture with the right hand for a blob rising up while the left hand for the bowling ball is floating motionlessly in the upper periphery

b. and as the bowling ball's coming down
Accompanied by an observer viewpoint iconic gesture with the left hand for the bowling ball coming down while the right hand for the character floats in the lower periphery

c. he swallows it
Accompanied by an observer viewpoint iconic gesture with the left hand (representing the bowling ball) passing inside the space formed by opening the right hand (representing the character's mouth).

The utterances of (15) explain how Tweety uses a bowling ball to frustrate Sylvester's attempt to climb up the drainpipe in the same cartoon described in (9). The three gestures exhibit the reuse of virtual space observed in (14). But in fact they go further: they offer similar forms in similar spatial configurations to represent the SAME objects over time. The right hand consistently represents Sylvester in all three utterances; the left hand consistently represents the bowling ball. This is a CATCHMENT in McNeill's (2000a, 2000b, 2005) terminology.

McNeill (2005) suggests that this persistent figuration offers evidence that the speech and gesture exhibited in (15) work together to form a coherent segment of the overall narrative. He also observes that (15) follows another segment of the discourse which provides the narrative background for its key events; the earlier segment explains how Sylvester climbs up the pipe and Tweety drops the ball. This earlier segment is delimited not only by its distinct content but by a distinct figuration in its gestures. For coherence theory, the CHANGE in figuration at the segment boundary beginning (15) is indicative of the interpretive preferences that always guide our resolution of underspecified meanings. In particular, what distinguishes the new gestures is that they depict BOTH Sylvester AND the bowling ball simultaneously. These entities are salient in previous discourse. In addition, both the previous discourse segment just concluded and the new discourse segment just begun focus on the INTERACTION between Sylvester and the bowling ball. This is a theme which dovetails with content that references both principals, which is exactly what the gestures of (15) offer.

Thus, this represents a case, in certain respects analogous to (10), where interpretation not only involves the reuse of salient information from the discourse context, but specifically involves the reuse of information at a level that engages with the overarching issue developed in this segment and the segment to which it attaches in discourse. Without such an assumption, it might be difficult to explain how the left hand in (15a) can depict a bowling ball when this entity is neither evoked in the associated speech nor inferable from it. Accordingly, the example suggests the theoretical possibilities of models of gesture interpretation that resolve underspecified information not just in light of previous discourse but in ways specifically guided by discourse structure—much as we find in the resolution of the underspecified meaning of verbal material. Of course, as we find parallels between the interpretation of gesture and that of speech—in abstract meaning, holistic resolution to salient information and respect for discourse structure—we bring into relief the problem of broadening coherence theory so as to characterise all aspects of gesture meaning and interpretation. This includes, for example, dimensions of iconicity and deixis that do not normally figure in the meaning of verbal material.

4 An Extended Example

With its emphasis on the relationships among communicative actions, and the role of discourse structure and discourse context in interpretation, coherence theory privileges the extended discourse as the object of semantic and pragmatic analysis. The examples we have discussed in Sections 2 and 3 suggest an analogous approach to the interpretation of gesture. But we have chosen our examples so far to cast particular elements of coherence theory into clear relief, rather than to show the interplay of these elements in richer and more complex cases. Ultimately, coherence theory explains speakers' orchestration of form and meaning in discourse in terms of the successive resolution of underspecified utterances against an evolving record of salient information and structure in the service of an improvised, interactive but concerted strategy to make a set of related ideas available for public discussion. Examples such as (16), as described below, highlight how these principles of coherence theory, as outlined in Sections 2 and 3, might jointly come to bear in typical cases of extended discourse. As such, they suggest the analytic power of coherence theory as a framework for investigating meaning and interpretation in embodied discourse.

We draw (16) from a collaborative learning session involving five adults studying physics.[1] The topic is Newton's law of gravity, which quantifies the force of gravitational attraction between interacting objects as a function of their masses and the distance between them. In discussion, the students observe that Newton's law gives rise to an apparent puzzle. According to Newton's law, objects of different masses experience different gravitational forces. Yet, qualitatively, Newton's law predicts that all objects move the same way under the influence of gravity. The students' discussion aims for a clear explanation that reconciles these two perspectives, in terms of the mathematical and physical relationships behind Newton's law.

[1] An extended recording of this interaction is available at www.talkbank.org/media/Class/Warren/gravity.mov. An extract highlighting the specific utterances we present in (16) can be viewed at homepages.inf.ed.ac.uk/alex/Gesture/gravity-eg.mov.

To ground the discussion, the students consider a specific case: the contrast between a mass of one unit and another mass of ten units. This specific case grounds their talk as they sharpen the problem and its solution. In describing the case to one another, the students not only converge on the specific words and concepts that define the problem—the concepts of mass, force, acceleration, and the ratio of one to ten units—they also arrive at a common figuration for illustrating their sample case on their hands. Two hands are held, side by side, shoulder height, modeling the two masses. The gesture seems to offer a visualisation of Galileo's famous experiment in which two balls of different weight are dropped at the same time from the same height to demonstrate that they fall at the same speed.

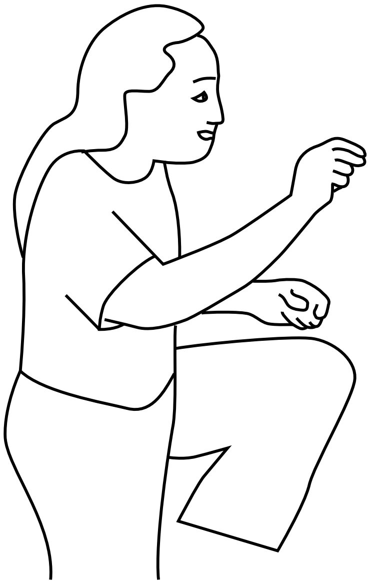
It is against this background that (16) occurs, and contributes the solution that the team eventually agrees on.

(16) a. If you see this larger ball as ten small balls like that.
The speaker gets down off his chair to match his interlocutor, Susan, who sits across from him on the floor. His right arm now extends out in front of him at shoulder height, with his fingers curled and his index finger touching his thumb, as though holding a pen (an ASL 1-flat handshape). During the gesture, the hand sweeps along a horizontal line further to the right.

b. They're all being pulled next to each other.
The speaker holds both hands at eye level, directly in front and above the shoulders, with elbows high and wide. The palms face slightly down, and the fingers are extended horizontally pointing at each other, demarcating a horizontal plane (like an ASL 5 flat handshape but with thumbs open).

c. Boom.
The speaker relaxes his elbows and hands, directing his fingers upwards and his palms to the sides, then sweeps his hands vertically downwards.

The explanation in (16) is preceded by another segment of interaction in which this speaker and another participant in the group session, Susan, interactively agree to consider a new thought experiment in the context of their ongoing problem solving. The dynamics of this segment are complex—they involve a disfluency on the part of the speaker, a subsequent repair effected simultaneously by the speaker and Susan, and overlapping gestures by the two interlocutors that track both the meanings they are working to convey and the processes of turn-taking that they are attempting to coordinate. We think—in keeping with the analysis we presented of (3), (4), and (6)—that such behaviour is compatible with coherence theory. However, since our main point in this paper is simply to justify the INFERENTIAL character of the interpretation in (16) itself, we merely attempt to summarise where this earlier segment leaves the interlocutors.



"If you see this ball as ten small balls like that"

Figure 1: Part of a Gestural Demonstration of a Galilean Experiment

The speaker of (16) has described the two balls which he will be dropping in his Galilean experiment. His accompanying gestures have acknowledged Susan for this setup, with a deictic orientation toward Susan herself, and toward the virtual space in front of her. Meanwhile, Susan has begun to echo the gestural setup of the experiment. She has raised her hands to shoulder height, holding a (virtual) small ball as a prop in her left hand, and holding her right hand with fingers loosely curled, palm facing right, evoking the action of holding a larger ball. Her continued bodily action as the explanation in (16) unfolds, we will see, demonstrates her evolving understanding of the speaker's words and gestures, and underscores the status of gestures in this interaction as central to the communication.

The speaker now carries out his modified thought experiment. Utterance (16a), as depicted on the right in Figure 1, presents the large ball as ten smaller balls, and indicates that an accompanying demonstration visualises the spatial relationships among these smaller balls, with the deictic *like that*. In tandem, the speaker appears to place a number of these imagined balls next to each other in a horizontal array in virtual space, as though ready to be dropped.

Utterance (16b) now demonstrates that each of the ten balls is subject to the same conditions. The speaker remarks on the common gravitational force that applies to the balls and, with *next to*, their common spatial alignment. The accompanying gesture underscores the precision of this alignment, as the plane of the speaker's fingers demarcates the calibrated starting level of the Galilean experiment. Already, by the end of (16b), the listener Susan drops her hands in anticipation of the conclusion of the experiment, and opening her mouth and raising her eyebrows in a display of astonished comprehension.

Finally, the speaker runs his experiment. The symmetry of the setup is clear—and has in fact been acknowledged by Susan, his primary interlocutor. So he concludes the explanation in extremely abbreviated form: the speaker simply announces the drop with *Boom* and dramatises on his hands the balls' synchronous fall.

Discourse (16) illustrates speakers' abilities to convey complex ideas effectively by deploying diverse elements of speech and gesture in strategic combination. Coherence theory offers a discipline to describe the inferential connections that hold such examples together and the inferential reasoning through which such examples are planned and understood. It thus suggests directions for more precise analyses both of the utterance itself, as a contribution to conversation, and of the processes of communication—such as grounding, acknowledging and agreeing—in which such complex utterances figure.

Consider (16a) to start. The gesture here stands in a privileged relationship to the words, in that the spatial layout implicitly demonstrated in the gesture serves as a referent for *that* in the modifier *like that*. However, this compositional relationship does not exhaust the relationship between gesture and speech in (16a). The gesture uses character viewpoint to depict an agent setting up a variant of the Galilean experiment, at the same time as the speech describes a perspective that we, as an audience, could take on that experiment. This relationship of CONTIGUITY, narrating overlapping events, reveals itself in the specialised interpretation we assign to both speech and gesture. For example, the performance of the gesture, by finishing its series of placements with a sweep to the right, gives the impression of an indefinite number of events. From the accompanying words, we know to understand it as a schematic depiction of ten of them. We also know that this depiction is understood conditionally, as part of the same hypothetical variant introduced in the associated words. Meanwhile, the words themselves do not explicitly evoke the Galilean figure, and might otherwise be understood as a more general exploration of the mathematics of Newton's laws. The gestures illustrate the specific context in which the perspective described by speech is to be taken.

This joint interpretation in fact represents a natural resolution of the more abstract meanings of the communicative actions involved that are revealed by just their form; it's natural partly because it uses salient information. For the gesture in particular, the contextual grounds for interpretation are clear. The Galilean experiment is the key case through which the interlocutors are working to resolve their puzzle about gravity, and the interlocutors have converged on a depiction of the Galilean experiment in the virtual space in front of their bodies. So it makes sense to understand the speaker's handshape and manner of motion here in terms of a character viewpoint depiction of the action of setting up balls to be dropped, and to understand the placements he indicates as recapitulating the virtual space in which the experiment unfolds.

The words and gesture of (16b), meanwhile, both portray CONSEQUENCES of the assumptions presented and depicted in (16a). Thus, the content conveyed in both speech and gesture is relativist to the conditional setting evoked in (16a), and the gesture in particular is understood to depict the same virtual space as the gesture of (16a). The speech and gesture of (16b) themselves seem to stand in a relationship of ELABORATION one to the other. The speech describes the balls as arrayed horizontally while the gesture suggests that the balls respect a measured equal height. Again, we see the interpretive effects of this coherence in the mutual disambiguation of speech and gesture, with *next to* understood not just in terms of *adjacency* but in terms of the horizontal array in which the balls are positioned, and the gesture understood with reference to the balls evoked as *they* in the associated utterance.

In addition, we can continue to see this interpretation not just as a coherent possibility, but as the specialisation of a more abstract meaning to salient information in the discourse context. In creating a flat surface with his extended fingers, the speaker marks a limiting boundary in virtual space, in a form similar to what one might use to demonstrate the height of a child—or, in a different orientation, the length of a fish. The abstract meaning of a demonstrated limit is resolved in the context of the Galilean experiment to the measured height from which objects will be dropped. In directing the fingertips of each hand toward those of the other, meanwhile, the speaker offers an image of precise alignment. The abstract matching indicated here again resolves to the understood setting of the Galilean experiment, where objects are positioned at a matching level to test whether they maintain it on their descent.

Finally, in (16c), the experiment unfolds. In context, we understand the sweep of the hands to model the descent of the array of balls, in synchrony with one another. We see the same virtual space and the same narrative continuation of the hypothetical experiment carried over here from (16a) and (16b). We understand *boom*, the underspecified description of a crashing sound, as the sound of the balls' synchronous impact with the ground. Gesture and speech are thus related in depicting the sound and motion of the balls' descent. Otherwise, though, there is little overlap in the content these actions offer. The descent is not even evoked explicitly in words.

If, as we have suggested here, coherence theory provides a discipline to capture the inferential connections that make (16) an effective presentation, with the consistent resolution of underspecified meaning drawn from salient information, then coherence theory also provides a discipline to describe the DIFFERENCES in pragmatic resources that characterise gesture as compared with speech. Thus, we find new relationships of DEPICTION, which show the distinctive kinds of information and the distinctive purposes which speaker's realise in their use of gesture to illustrate associated verbal material. And we find new relationships BETWEEN gestures, as in the OVERLAY relationship through which successive gestures reference a common virtual space. For as McCullough (2005) attests, gestures often convey spatial information that is absent from the narrative in the speech. Such relationships call for an extension of coherence theory, challenging us to see coherence not just as a property of the use of words but more generally as a property of any deliberate presentation of content: for instance, analysing gesture requires the introduction of coherence relations whose semantic entailments constrain the spatial relations among individuals.

We also find new kinds of configurations as we organise these interpretive connections to arrive at the structure of a coherent discourse. It is tempting, for example, to characterise the structural relationship between (16a) and (16b) in terms of coherence relationships connecting successive verbal phrases (RESULT, here), coherence relationships connecting successive gestures (OVERLAY, here), coherence relationships connecting associated speech and gesture into utterances (different cases of CONTIGUITY and DEPICTION, here), and coherence relationships connecting the successive multimodal utterances themselves (HYPOTHESIS–CONCLUSION here). In fact, we might ultimately want to fit the interlocutor Susan's mirroring gestures within the same interpretive structure (Lascarides & Asher, 2009). The interaction of speech and gesture in this case invites us to adopt a perspective on discourse structure not just as a tree describing the relationship of successive clauses but as a systematically layered framework befitting a orchestrated program of synchronous communicative action.

Similarly, to say that gesture meaning, like word meaning, must be described in terms of abstract constraints resolved to salient values in context, is not to say that gestures carry the same kinds of meanings as words, or are resolved against the same information. We need only observe how the gestures of (16) reference bodily action in physical space, through physically-grounded concepts of placement, measurement, and motion, and are resolved against a correspondingly concrete image of a person physically setting up and carrying out a Galilean experiment. Analysts would have little reason to postulate such abstract meanings or such concrete instances in this context on the basis of the interpretations of the words alone. One might nevertheless suspect that this embodied dimension of understanding is always present in human thought and communication (Lakoff & Johnson, 1999). If so, the developments of coherence theory that are required to accommodate these new dimensions of meaning and context to interpret gesture can also help us to articulate more precisely the speaker's meaningful engagement with his audience, but also give a unique insight into the thinking that the speaker is working to share.

5 Other Approaches

Theories of coherence respond to many of the same phenomena and desiderata as other approaches to pragmatic interpretation in linguistics and philosophy, from relevance theory (Sperber & Wilson, 1986) to dynamic semantics (Beaver, 2001). The commonality suggests that any of these approaches might find evidence in the examples of Sections 2, 3 and 4 to extend their principles and mechanisms to the integrated interpretation of speech and gesture. We are unaware of attempts to do this, however.

In any case, what distinguishes coherence from these approaches is its methodological commitment to articulating explicit computational algorithms and representations for reconstructing the preferred interpretation in context. Such algorithms differ in the weight they give to linguistic knowledge and communicative conventions on the one hand, and real-world knowledge and principles of rationality on the other. At one extreme, the early work of Grosz and Sidner (1986, 1990) and later work of Lochbaum (1998) epitomise the possibility of describing discourse coherence through general principles of agent rationality, abstracting away almost completely from the words and meanings of linguistic utterances. At another extreme, the work of Asher and Lascarides (2003) epitomises the possibility of characterising coherence largely through conventional default inferences defined directly over the semantic representations delivered by a formal grammar, while factoring out cognitive modeling to a large extent. In between fall approaches like the theory of interpretation as abduction explored by Hobbs et al (1993) and Kehler (2002), which subject linguistic representations directly to general purpose inference. Despite these differences, the principles and insights of the approaches are largely complementary. The differences may have more to do with the philosophical perspective that researchers use to narrate their formal devices than with the substantive differences in the mechanisms and predictions of the different approaches (Stone, 2004b, 2004a). We have largely abstracted away from computational issues in this paper. A first stab at a formal account of the interpretation of language and gesture, within the framework of Segmented Discourse Representation Theory (SDRT) developed by Asher and Lascarides (2003), can be found elsewhere (Lascarides & Stone, 2006).

With this in mind, we contrast our approach specifically with other computational approaches to the synthesis of gesture in association with speech (Cassell, Stone, et al., 1994; Cassell, Stone, & Yan, 2000; Kopp et al., 2004) and its recognition (Johnston et al., 1997; Johnston, 1998). This work has largely focused on a different problem to ours. All these researchers are primarily concerned to capture the synchrony between speech and gesture in a suitable representation of utterance structure, and thereby to use the mechanisms of compositional syntax and semantics to identify the units of interpretive interaction in complex multimodal utterances. We are in agreement that such structures and mechanisms are implicated in utterance interpretation. What we argue is that coherence is also necessary, to describe the reasoning that guides interpretation, and its results.

Cassell and colleagues have characterised the interpretive relationship between gesture and speech in terms of a distinguished shared entity reference that controls the synthesis of an appropriate gesture in tandem with speech. Both the gesture and its associated speech must supply content that describes this entity. The value of this entity is determined by the compositional structure and discourse function of the verbal material (Cassell, Pelachaud, et al., 1994). What synchronises with gesture must be a RHEME, that is, verbal material that provides the main contribution of an utterance by addressing a salient open question in the discourse. Other verbal material constitutes the THEME, which identifies the open question being addressed. An initial implementation synthesised gestures by retrieving a standardised depiction of the entity described by the rheme (Cassell, Pelachaud, et al., 1994). Later systems selected gestures in context so as to help fulfill a specified set of communicative goals that could be realised either in gesture or in speech (Cassell et al., 2000). The system represented the input communicative goals and the output content in an utterance as a flat list, however, so that coherence between speech and gesture was not an explicit theoretical construct, but arose only implicitly in the system through the process of discourse planning. (Kopp et al., 2004) improve the model further by synthesising gestures based on multiple dimensions of form and meaning, through a model broadly consonant with the multidimensional characterisation presented for gestures like (14) in Section 3. However, this model also leaves coherence implicit, and continues to use shared entity reference to characterise the co-expressiveness of speech and gesture.

In comparison to these models, coherence theory suggests a looser relationship between what gesture depicts and the entities referenced in speech. The examples of Section 2, among others, show that gestures do not always share entity references with the associated speech. At the same time, coherence theory suggests a TIGHTER relationship between the content of gesture and that of speech. The examples of Section 3, among others, show that such interpretive relationships are needed in the general case to resolve the underspecified meanings of gestures in specific contexts.

Johnston and colleagues (1997; 1998) offer a computational characterisation of the use of pen gestures in association with spoken instructions to user interfaces. While pen gestures obviously differ in form from hand gestures, both modalities exhibit deictic and iconic meanings and a close interrelationship with associated speech. Johnston develops grammar rules that specify the joint interpretation of complex utterances, as a function of the interpretations of the actions across modalities that make them up, and the way those actions are coordinated. Johnston's approach aligns with coherence theory in that the relationship in reference between speech and gesture may be mediated by an indirect semantic relationship—albeit one specified in the grammar rather than derived by contextual inference. Johnston's approach also offers a holistic process of interpretation, searching to construct an overarching analysis of the utterance in terms of grammatical rules, and disambiguating fragmentary features of gesture in the process. However, integrated interpretation is limited to selecting alternative analyses rather than resolving underspecification through inferred links to context. Thus, it is much more natural to use coherence theory to describe gestures such as (14), (15), or (16b), which exhibit a multifaceted interpretation with an interpretive relationship to previous gesture and inferential access to commonsense background knowledge, as well as a specific relationship with the associated speech.

A parallel approach to ours has been pursued independently by (Lücking et al., 2006), who focus on situated utterances that combine language with physical action in the world, such as steps of assembly. Their data, account and arguments are quite different to ours, focusing on the resolution of fragmentary utterances, particularly deictic references. Nevertheless, they explain their results by appeal to a framework broadly similar to ours. While our insights are broadly compatible, (Lücking et al., 2006) in fact consider only deictic gestures, and model their content as a single entity reference, linked with the associated speech through grammatical means, as proposed by (Rieser, 2005). Thus they do not explore the CONTENT of gesture, or the role of coherence in resolving that content and linking it to the discourse context.

6 Conclusion

People's contributions to face-to-face conversation are not just words. As cases like (16) make clear, they are sophisticated orchestrations of linguistic utterances with expressive movements. In such cases, interlocutors offer and understand visual explanations of their ideas.

Gestures get their meanings in very different ways to words. But both words and movements have meanings that are ambiguous and underspecified, and need to be resolved in context. When we make sense of a visual explanation, as in (16), we succeed in resolving these gaps and knitting what we see and hear into an understanding of our interlocutor, and of the complex idea that they hope and strive to share with us.

In this paper, we have explored and argued for an approach to this sense-making based on theories of coherence from computational semantics and pragmatics. These theories allow us to capture how interlocutors' understanding of one another's communicative intentions originates in general knowledge of meaning and in interpretive principles for finding coherence. In particular, we have argued that we work to construct coherent interpretations by identifying semantic relationships that group resolved meanings together into hierarchical structures that explain what the speaker is doing in an attempt to communicate with us.

Plainly, this is the kind of work that sets an agenda rather than solves a problem. The challenge now is to formalise communicative actions across modalities through general models of coherence, and to support those formal models with general arguments. In particular, our arguments in this paper have drawn on descriptions of specific attested examples. The linguistics literature, by contrast, often motivates its accounts by considering informants' judgements about constructed utterances. Minimal pairs—utterances that differ just in a single dimension of analysis—have proved particularly crucial for linguistic argumentation. The complex structure, timing and realisation of embodied utterances, however, makes it much more problematic to present minimal pairs and to characterise their interpretations. To develop general argumentation may therefore require new methodologies. One possibility would be to combine utterance synthesis via animation with ways to elicit reliable judgements about the resulting performances, extending the preliminary work of (Kopp et al., 2004; Stone et al., 2004; Stone & Oh, 2008). This formal development is just beginning (Lascarides & Stone, 2006). It will require the same diverse and sustained effort that has been and remains underway in the formalisation and implementation of coherence for purely linguistic discourse, as we confront and engage with the many difficulties of generalising coherence theories to this broader communicative arena. We hope the analyses and arguments we have presented showcase the value and need for such efforts.

Acknowledgements

This work has been presented at several workshops and conferences: the Workshop on Embodied Communication (Bielefeld, March 2006); Constraints in Discourse (Maynooth, June 2006); the Pragmatics Workshop (Cambridge, September 2006), Brandial (Potsdam, September 2006); the Workshop on Dynamic Semantics (Oslo, September 2006); Dialogue Matters (London, February 2008); the Rutgers Semantics Workshop (New Jersey, November 2008). We would like to thank the participants of these for their very useful comments and feedback. Finally, we would like to thank the many individuals who have influenced this work through discussion and feedback: Nicholas Asher, Susan Brennan and her colleagues at Stony Brook, Justine Cassell, Herb Clark, Susan Duncan, Jacob Eisenstein, Gabriel Greenberg, Jerry Hobbs, Michael Johnston, Hannes Rieser and his colleagues at Bielefeld, Candy Sidner, and Rich Thomason. We also owe a special thanks to Jean Carletta, who provided support in our search of the AMI corpus, including writing search scripts that helped us to find what we were looking for. Finally, we would like to thank two anonymous reviewers for this journal, and also the editor Adam Kendon, for their very detailed and thoughtful comments on earlier drafts of this paper. Any mistakes that remain are our own.

Much of this research was done while Matthew Stone held a fellowship in Edinburgh, funded by the Leverhulme Trust. We would also like to thank the following grants from the National Science Foundation (NSF) for their financial support: HLC-0308121, CCF-0541185, HSD-0624191.

The Authors

Alex Lascarides is a Reader in the School of Informatics at the University of Edinburgh. She received her PhD in Cognitive Science from the University of Edinburgh. Her research interests are in formal and computational semantics and pragmatics, and together with Nicholas Asher she is the principal developer of Segmented Discourse Representation Theory. She was chair of the European Association for Computational Linguistics from 2007–2008.

Matthew Stone is Associate Professor in the Computer Science Department and the Center for Cognitive Science at Rutgers. He got his PhD in 1998 from the University of Pennsylvania. He studies computational models of conversation, particularly models of utterance production, for intelligent agents that interact naturally with human partners. He recently concluded a term on the editorial board of the journal Computational Linguistics and served as program co-chair for the 2007 North American Association for Computational Linguistics Human Language Technology Conference (NAACL HLT).

References

Asher, N. (1993). *Reference to abstract objects in discourse*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Asher, N., & Lascarides, A. (2003). *Logics of conversation*. Cambridge, UK: Cambridge University Press.

Bavelas, J. B., & Chovil, N. (2000). Visible acts of meaning: An integrated message model of language in face-to-face dialogue. *Journal of Language and Social Psychology*, *19*(2), 163–194.

Beaver, D. (2001). *Presupposition and assertion in dynamic semantics*. Stanford, USA: CSLI. (Revision of 1995 PhD Thesis, University Of Edinburgh)

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology*, *22*(6), 1482–1493.

Cassell, J. (2001). Embodied conversational agents: Representation and intelligence in user interface. *AI Magazine*, *22*(3), 67–83.

Cassell, J., McNeill, D., & McCullough, K.-E. (1999). Speech-gesture mismatches: evidence for one underlying representation of linguistic and nonlinguistic information. *Pragmatics and Cognition*, *7*(1), 1–33.

Cassell, J., Pelachaud, C., Badler, N., Steedman, M., Achorn, B., Becket, T., et al. (1994). Animated conversation: Rule-based generation of facial expression, gesture and spoken intonation for multiple conversational agents. In *Proceedings of siggraph* (pp. 413–420). Orlando.

Cassell, J., Stone, M., Douville, B., Prevost, S., Achorn, B., Steedman, M., et al. (1994). Modeling the interaction between speech and gesture. In *Proceedings of the cognitive science society.* Atlanta.

Cassell, J., Stone, M., & Yan, H. (2000). Coordination and context-dependence in the generation of embodied conversation. In *First international conference on natural language generation* (pp. 171–178). Mitze Ramon, Israel.

Emmorey, K., Tversky, B., & Taylor, H. (2000). Using space to describe space: Perspective in speech, sign and gesture. *Spatial Cognition and Computation*, *2*(3), 157–180.

Engle, R. (2000). *Toward a theory of multimodal communication: Combining speech, gestures, diagrams and demonstrations in structural explanations*. Unpublished doctoral dissertation, Stanford University.

Ginzburg, J., Fernandez, R., & Schlangen, D. (2007). Unifying self- and other-repair. In *Proceedings of the 2007 workshop on the semantics and pragmatics of dialogue (decalog).* Trento, Italy.

Grosz, B., & Sidner, C. (1986). Attention, intentions and the structure of discourse. *Computational Linguistics*, *12*, 175–204.

Grosz, B., & Sidner, C. (1990). Plans for discourse. In J. M. P. R. Cohen & M. Pollack (Eds.), *Intentions in communication* (pp. 365–388). Cambridge Mass.: MIT Press.

Haviland, J. B. (2000). Pointing, gesture spaces, and mental maps. In D. McNeill (Ed.), *Language and gesture* (pp. 13–46). Cambridge, UK: Cambridge University Press.

Heeman, P., & Hirst, G. (1995). Collaborating on referring expressions. *Computational Linguistics*, *21*(3), 351–382.

Heeman, P. A., & Allen, J. F. (1999). Speech repairs, intonational phrases, and discourse markers: modeling speakers' utterances in spoken dialogue. *Computational Linguistics*, *25*(4), 527–571.

Hobbs, J., Stickel, M., Appelt, D., & Martin, P. (1993). Interpretation as abduction. *Artificial Intelligence*, *63*(1–2), 69–142.

Hobbs, J. R. (1979). Coherence and coreference. *Cognitive Science*, *3*(1), 67–90.

Horn, L. R. (1991). Given as new: When redundant affirmation isn't. *Journal of Pragmatics*, *15*, 305–328.

Johnston, M. (1998). Unification-based multimodal parsing. In *Coling/acl.* Montreal, Canada.

Johnston, M., Cohen, P., McGee, D., Pittman, J., Oviatt, S., & Smith, I. (1997). Unification-based multimodal integration. In *ACL/EACL 97: Proceedings of the annual meeting of the assocation for computational linguistics.* Madrid, Spain.

Kehler, A. (2002). *Coherence, reference and the theory of grammar.* Stanford, USA: CSLI Publications, Cambridge University Press.

Kendon, A. (2004). *Gesture: Visible action as utterance*. New York, USA: Cambridge University Press.

Kopp, S., Tepper, P., & Cassell, J. (2004). Towards integrated microplanning of language and iconic gesture for multimodal output. In *Proceedings of icmi.* Pennsylvania.

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

Krauss, R. M., Chen, Y., & Gottesman, R. F. (2000). Lexical gestures and lexical access: a process model. In D. McNeill (Ed.), *Language and gesture* (pp. 261–283). New York: Cambridge.

Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to western thought*. New York: Basic Books.

Lascarides, A., & Asher, N. (1993). Temporal interpretation, discourse relations and commonsense entailment. *Linguistics and Philosophy*, *16*(5), 437–493.

Lascarides, A., & Asher, N. (2009). Agreement, disputes and commitment in dialogue. *Journal of Semantics*, *26*(2).

Lascarides, A., & Stone, M. (2006). Formal semantics for iconic gesture. In *Proceedings of the 10th workshop on the semantics and pragmatics of dialogue (brandial).* Potsdam.

Levy, F. N. (Ed.). (1913). *American art annual* (Vol. X). New York: American Federation of Arts.

Lewis, D. (1979). Scorekeeping in a language game. *Journal of Philosophical Logic*, *8*, 339–359.

Lochbaum, K. (1998). A collaborative planning model of intentional structure. *Computational Linguistics*, *24*(4), 525–572.

Lücking, A., Rieser, H., & Staudacher, M. (2006). SDRT and multi-modal situated communication. In *Proceedings of BRANDIAL.* Potsdam.

Macaulay, D. (1988). *The way things work.* Boston, Mass.: Houghton Mifflin.

Mann, W. C., & Thompson, S. (1987). Rhetorical structure theory: A framework for the analysis of texts. *International Pragmatics Association Papers in Pragmatics*, *1*, 79–105.

McCullough, K.-E. (2005). *Using gfestures in speaking: Self-generating indexical fields*. Unpublished doctoral dissertation, University of Chicago.

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

McNeill, D. (2000a). Catchments and context: Non-modular factors in speech and gesture. In D. McNeill (Ed.), *Language and gesture* (pp. 312–328). New York: Cambridge University Press.

McNeill, D. (2000b). Growth points, catchments and contexts. *Japanese Journal of Cognitive Science*, *7*(1), 22–36.

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

McNeill, D., Quek, F., McCullough, K.-E., Duncan, S., Furuyama, N., Bryll, R., et al. (2001). Catchments, prosody and discourse. *Gesture*, *1*, 9–33.

Moens, M., & Steedman, M. (1988). Temporal ontology and temporal reference. *Computational Linguistics*, *14*(2), 15–28.

Özyürek, A. (2001). What do speech–gesture mismatchhes reveal about language specific processing? a comparison of turkish and english. In *Proceedings of the 27th meeting of the berkeley linguistics society.* Berkeley.

Poesio, M., & Traum, D. (1997). Conversaiontal actions and discourse situations. *Computational Intelligence*, *13*(3).

Rieser, H. (2005). Pointing and grasping in concert. In M. Stede, C. Chiarcos, M. Grabski, & L. Lagerwerf (Eds.), *Salience in discourse: multidisciplinary approaches to discourse* (pp. 129–139). Münster: Nodus.

Schlangen, D. (2003). *A coherence-based approach to the interpretation of non-sentential utterances in dialogue*. Unpublished doctoral dissertation, University of Edinburgh.

Sperber, D., & Wilson, D. (1986). *Relevance*. Oxford: Blackwells.

Stone, M. (2004a). Communicative intentions and conversational processes in human-human and human-computer dialogue. In J. Trueswell & M. K. Tanenhaus (Eds.), *World-situated language use: Psycholinguistc, linguistic and computational perspectives on bridging the product and action traditions.* Cambridge, Mass.: MIT.

Stone, M. (2004b). Intention, interpretation and the computational structure of language. *Cognitive Science*, *28*(5), 781–809.

Stone, M., DeCarlo, D., Oh, I., Rodriguez, C., Stere, A., Lees, A., et al. (2004). Speaking with hands: Creating animated conversational characters from recordings of human performance. *ACM Transactions on Graphics*, *23*(3), 506–513.

Stone, M., Doran, C., Webber, B., Bleam, T., & Palmer, M. (2003). Microplanning with communicative intentions: The SPUD system. *Computational Intelligence*, *19*(4), 311–381.

Stone, M., & Oh, I. (2008). Modeling facial expression of uncertainty in conversational animation. In I. Wachsmuth & G. Knoblich (Eds.), *Modeling communication with robots and virtual humans* (pp. 57–76). New York: Springer.

Strawson, P. (1952). *Introduction to logical theory*. London: Methuen.

Thomason, R. (1990). Accommodation, meaning, and implicature: Interdisciplinary foundations for pragmatics. In P. R. Cohen, J. Morgan, & M. Pollack (Eds.), *Intentions in communication* (pp. 325–363). Cambridge, Massachusetts: MIT Press.

USAF. (1988). *Organizational maintenance job guide (fuel system distribution, USAF series*

F-16C/D aircraft). United States Air Force.

Walker, M. A. (1993). *Informational redundancy and resource bounds in dialogue*. Unpublished doctoral dissertation, Department of Computer & Information Science, University of Pennsylvania. (Institute for Research in Cognitive Science report IRCS-93-45)

Webber, B. (1988). Tense as discourse anaphor. *Computational Linguistics*, *14*(2), 61-73.

Webber, B. (1991). Structure and ostension in the interpretation of discourse deixis. *Natural Language and Cognitive Processes*, *6*(2), 107–135.

Webber, B., Knott, A., Stone, M., & Joshi, A. (2003). Anaphora and discourse structure. *Computational Linguistics*, *29*(4), 545–588.