

Interactive Context in Sketch Understanding

Patrick G.T. Healey
Information, Media and Communication
Research Group,
Department of Computer Science,
Queen Mary University of London.

**Simon Garrod &
Nick Fay**
HCRC,
Department of Psychology,
University of Glasgow.

**John Lee &
Jon Oberlander**
HCRC,
Division of Informatics
University of Edinburgh.

Abstract

A substantial body of empirical evidence indicates that interactional context has a key influence on the form and interpretation of language. This paper provides an overview of a series of experiments which indicate that interactional context also plays a key role in the interpretation of drawings and sketches. Two experimental graphical communication tasks, analogous to definite reference tasks, are described. The findings from these tasks show significant parallels between the mechanisms of co-ordination in graphical dialogue and natural language dialogue. Specifically; participants match on drawing types above chance, 'graphical referring expressions' reduce with repetition in an interaction, direct interaction promotes the use of more abstract drawings, and community-specific graphical conventions emerge in experimental 'sub-communities'.

Interactive Context in Dialogue

Conversation is a, if not *the*, key context of understanding for language. People's use of language to represent objects, events and situations is sensitive to, amongst other things; who they are speaking to, the mutual availability of referents, the history of their conversation and their (dis)joint membership of cultural and linguistic sub-communities (Hymes 1972; Clark 1998). Evidence for the direct influence of interactional context on interpretation and understanding comes from a variety of sources (see Krauss & Fussell, 1996, for a review). One example of the influence of interactional context on interpretation is provided by work on the Collaborative Model of dialogue. Schober & Clark (1989), Clark & Wilkes-Gibbs (1986), and Wilkes-Gibbs & Clark (1992) have shown that full understanding of referring expressions depends on the degree of active participation in conversation by speaker and addressees. Non-active participants in a conversation, such as passive side-participants, overhearers, or bystanders, are less able to understand referring expressions than active participants. This may occur even when, in simple informational terms, they are equivalent to active participants.

A second example of the influence of interactional context comes from studies of conceptual and linguistic co-ordination in dialogue. Anderson & Garrod (1987), Garrod

& Anderson (1987) have shown that conversational partners show a marked tendency to match or 'entrain' the form and interpretation of utterances during interaction. For example, where several types of semantically distinct referring expressions are possible for describing a location, people show a strong preference for matching the type of expression used by their conversational partner (Garrod & Anderson 1987). Brannigan, Pickering, & Cleland (2000) have observed similar entrainment effects at the level of syntax. Garrod & Anderson (1987) argue that these dialogue phenomena reflect the operation of a basic dialogue co-ordination mechanism which simplifies the production and interpretation of utterances in interaction.

Intuitively, it might be supposed that graphical representations would be less sensitive to interactional context. One reason for this is that the production and use of drawings and sketches is normally treated, and analysed, as an activity more akin to monologue than dialogue (Scaife & Rogers, 1996). There is evidence, however, that this underestimates both the actual and potential use of drawing activities as a mode of interaction. Anecdotally, drawings are often incrementally produced and modified as part of a conversational exchange. For example, sketch maps and explanatory diagrams form a familiar extension of many routine conversations.

van Sommers (1984) provides evidence from a questionnaire study that approximately half of routine, non-work, drawing activities take place with or for an audience. Although van Sommers does not report how often these exchanges involve direct graphical interaction, his findings demonstrate the variety of interactional contexts in which drawing occurs. The most frequently cited category is the production of sketch maps of a local area, either as part of an explanation or in order to give directions. The second most frequently cited category relates to activities with children including; games and amusements, teaching or helping with homework and helping children learn to draw. Additional categories of 'public' drawing include; sketching of hair, makeup and clothing, sketching house plans, drawing to express feelings, defacing pictures and drawing people.

The collaborative development and modification of sketches is a feature of many specialised work related interactions, such as architect-architect and architect-client (Neilsen & Lee 1994; Healey & Peters 2001). We estimate

that in the architects' practice studied by Healey & Peters (2001) approximately 30% of drawing activities occurred as an integrated part of a conversational exchange. Engle (1998) provides experimental evidence that graphics, gesture and language combine in explanatory dialogues to create composite communicative signals (see also Clark, 1996). Overall, there is a clear *prima facie* case that sketches and drawings are often closely integrated into interaction and that this may have significant implications for their interpretation.

A second possible reason for scepticism about the role of interactional context in the interpretation of sketches and drawings is the intuition that drawings and sketches are easier to interpret than language. Arguably, many of the interactional influences on language interpretation are associated with the conventional nature of linguistic representation. Co-ordinated interpretation of utterances requires the concerted application of conventions. Interaction is used to maintain and modify those interpretations. Drawings and sketches can exploit iconicity to provide a less arbitrary form of representation. Consequently, we might suppose that they would be less dependent on interaction to secure their interpretation. While this might be true in cases such as sketches of buildings or people, it does not cover the range of uses to which sketches and drawings are put. Explanations involving sketches of Venn diagrams or Euler circles provide perhaps the most obvious counter-example.

The present paper summarises the findings from a series of experiments, carried out as part of the MAGIC project.¹ These experiments provide evidence that the interpretation of drawings and sketches is sensitive to interactional context. In particular, that interactional context has marked effects on the form, interpretation and understanding of sketches; and that the mechanisms and processes that give rise to these effects show substantial parallels to those identified for natural language dialogue.

Interactional Context in Graphical Dialogue

The experiments involve two basic referential communication tasks, the Concept Drawing task and the Music drawing task, in which pairs of subjects communicate about a variety of concepts using exclusively graphical means. These tasks can be thought of as two-way or conversational variants of the party game Pictionary.

¹The MAGIC -Multimodality and Graphics in Interactive Communication- project is a collaboration between the Information, Media and Communication Research group in the Department of Computer Science at Queen Mary University of London, the Human Communication Research Centre and Department of Architecture at the Universities of Edinburgh and Glasgow and ATR International Media Information Sciences Laboratory, Kyoto (<http://www.dcs.qmul.ac.uk/research/magic/magic.html>). It is funded by under the ESRC/EPSC PACCIT programme: People at the Centre of Computers and Information Technology, grant number L328253003. We also gratefully acknowledge James King, Nik Swoboda, Ichiro Umata and Yasuhiro Katagiri who contributed to the experiments summarised in this paper.

The Experimental Tasks

In the basic Concept Drawing Task, an ordered list is produced consisting of twelve concept words drawn from the categories; places (e.g., "theatre", "art gallery", "museum"), people (e.g., "Robert de Niro", "Arnold Schwarzeneger", "Clint Eastwood"), television programmes (e.g., "drama", "soap-opera", "cartoon"), objects (e.g., "television", "computer microwave"), and abstract concepts (e.g., "loud", "homesick", "poverty"). One participant, the 'Drawer', is provided with the ordered list of twelve words. Their partner, the 'Chooser', is presented with an unordered list including four extra distractors i.e., a total of sixteen words. The task is for the Drawer to take each word in turn and produce a sketch of it so that their partner, the Chooser, can identify the item in their list. The aim is for the Chooser to determine the original list of twelve items that the Drawer started with.

The basic Music Drawing Task is similar to the Concept Drawing task however, pieces of music are used in place of concept words. The pieces are all thirty second piano solos taken from a variety of styles. In order to avoid any verbal mediating strategies, only pieces whose titles are relatively unknown are used. The typical procedure is that the Drawer listens to a target piece of piano music and produces a sketch of it. The Chooser has two pieces of music, the target and a distractor, and tries to select which is the piece depicted by the Drawer. To avoid overhearing of the original pieces the Drawer and Chooser are seated in separate rooms. Playback of the pieces is self-paced and all drawing takes place on a shared virtual whiteboard.

In both tasks, subjects are free to use any style of drawing they like; the only restriction on drawing is that they do not use letters or numbers. The types of drawing produced for each concept or piece of music varies substantially between pairs, some examples are provided in Figures 1, 2 and 3. All things being equal, each pair tends to develop their own conventional solutions to the communication problems posed by the task. Subjects find the tasks enjoyable and engaging and perform them with above chance accuracy. Drawings have been carried out on both conventional physical whiteboards and virtual whiteboard software developed specifically for the MAGIC project (Healey, Swoboda, & King 2002).

Effects of Interactional Context

A number of experiments have been performed using these tasks which suggest important parallels between the effects of interactional context on graphical and linguistic dialogue. Here we provide an overview of the main findings from these experiments and explore some of their implications for sketch understanding.

Interactional Entrainment. One of the simplest pieces of evidence for effects of interactional context on the use of drawing comes from the Music Drawing task. Participants in this task produce drawings that can be reliably classified into two basic types; 'Abstract' and 'Figurative' (Interjudge agreement: Kappa = 0.9, N = 287, k = 2). Abstract drawings, illustrated in Figure 1, typically involve graph-like representations of e.g., pitch, melody, rhythm or intensity. By

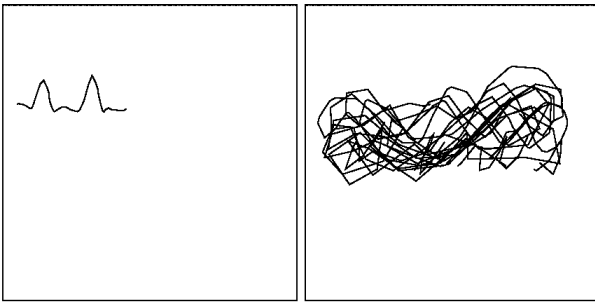


Figure 1: Example Abstract drawings from Two Successive Trials of the Music Drawing Task

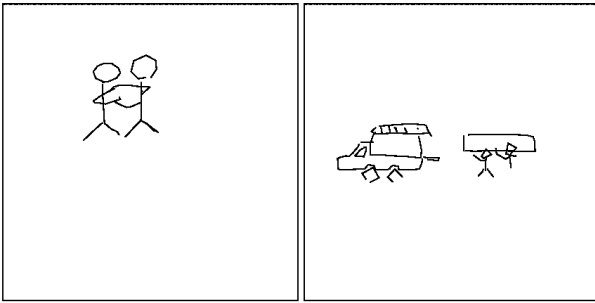


Figure 2: Example Figurative Drawings from Two Successive Trials of the Music Drawing Task

contrast, Figurative drawings, illustrated in Figure 2 typically depict recognisable objects, figures, landscapes or situations. Where pairs of participants in the task both take the role of Drawer, either by alternating roles or in manipulations in which both participants draw at the same time, they show a reliable tendency to match in their use of drawing the Figurative and Abstract drawing types (Healey *et al.* 2001). As noted in the introduction, this pattern of entrainment between the participants in an interaction is also well established for natural language dialogue at the levels of semantics and syntax (Garrod & Anderson 1987; Brannigan, Pickering, & Cleland 2000). Garrod & Anderson (1987) argue that entrainment constitutes a basic mechanism through which conceptual co-ordination is achieved in dialogue.

Contraction of Recurrent References. The procedure for the Concept Drawing task typically requires pairs to repeat the same set of twelve target words, in different orders, over several trials. This manipulation ensures that each word is drawn, and identified, several times by each pair. This is designed to reproduce the experimental procedure followed by Clark & Wilkes-Gibbs (1986) who investigated the production of recurrent (verbal) referring expressions by conversational partners. Clark and Wilkes-Gibbs found that both the average number of words and average number of turns used to refer to a target item (in their case a tangram figure) rapidly declined with the number of repetitions. Experiments with the concept drawing task show the same pattern

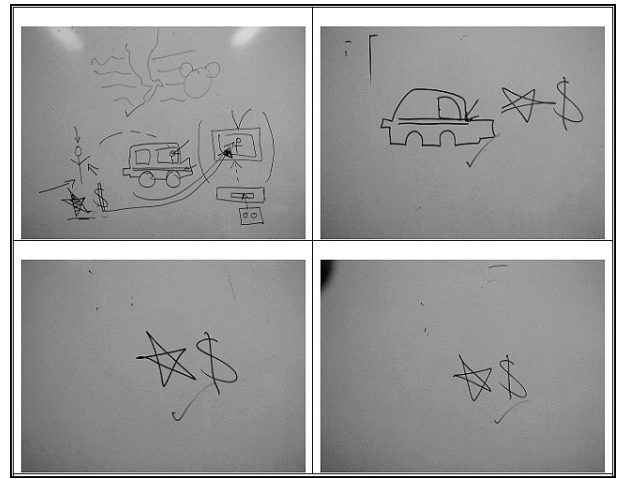


Figure 3: A Sequence of ‘Robert deNiro’s from the Concept Drawing Task

of reduction. This is illustrated in Figure 3 which shows a sequence of four trials (ordered left to right and top to bottom). Where target concepts recur, the drawings that represent them quickly become simplified. This is indexed both by simple quantitative measures such as the amount of ‘ink’ and number of lines used, their complexity as estimated by human judges and calculations of their visual complexity.²

Experiments to evaluate the effects of these contractions on the interpretability of the drawings by third parties are currently in progress. However, it appears likely that these reductions will have a substantial effect on recognisability. The first drawing in Figure 3 has a number of elements that might allow a non-participating observer to guess the identity of the individual depicted. For example it includes a sketch map of Italy, sketches of a TV and VCR, and an image of a taxi (which refers to a de Niro film). However the last sketch in the sequence, consisting of a star and a dollar sign would be much harder to decipher.

Effects of Direct Interaction. Experiments with the Music Drawing task have investigated the influence of level of communicative interaction between participants on the type of drawing (Abstract or Figurative) that they produce. The basic contrast is between an interactive and non-interactive version of the task (Healey *et al.* 2001; forthcoming). In the non-interactive version, subjects alternate between acting as Drawer and Chooser on each trial and only the Drawer can draw on each trial. In this version of the task each trial can be thought of as a single turn in the communicative exchange. In the interactive version the task is altered so that both members of a pair draw at the same time. They have one piece of music each and they must determine, using only drawing, whether their pieces are the same or different. In this version of the task the commu-

²The analysis of visual complexity is based on a psychophysical measure developed by Pelli *et al.* (in press) and is based on the formula: Complexity = Perimeter² / Ink.

nicative exchange is much richer. In addition to producing drawings of their pieces, subjects employ devices such as arrows, underlining, and circling to query and revise various aspects of their drawings. Each trial in the interactive task thus contains a number of ‘conversational’ turns.

The effect of the difference in level of communicative interaction can be seen in Table 1 (the ‘Composite’ category refers to drawings that combine Figurative and Abstract elements). Where both members of a pair can interact directly on the whiteboard, they rely primarily on the Abstract drawings. In the non-interactive task, where they are alternating between drawing and choosing, they rely primarily on Figurative drawings.

Table 1: Distribution of Drawing Types in the Music Drawing Task

Task Version	Drawing Type		
	Abstract	Figurative	Composite
Interactive	59%	21%	16%
Non-Interactive	27%	64%	8%

Healey *et al.* (forthcoming) also present evidence from the interactive version of the Music Drawing task that the difference in patterns of use of the drawing types is not explained by appeal to differences in the efficiency or effectiveness of the Abstract and Figurative drawing types. These two categories are not distinguishable in terms of the number of lines or ink (pixels) involved in producing them, nor in terms of the accuracy of responses associated with drawings of each type. Rather, it appears that it is the availability of specific mechanisms of communicative interaction, such as the circling and underlining of both their own *and* their partners drawing, that is critical. This claim is supported by the finding that subjects drawing activities overlap approximately 20% more when they produce Abstract drawings than when they produce Figurative drawings. Considered together these findings indicate that direct interaction is critical to the co-ordinated use of the Abstract drawings.

Community-based Conventions. Perhaps the most interesting evidence of effects of interactional context on sketch interpretation comes from experiments on the emergence of graphical conventions in experimental ‘sub-communities’ (cf. Garrod & Doherty, 1994).

Data from an unpublished experiment with the Music Drawing task demonstrates that, for the Music Drawing task at least, the patterns of co-ordination in drawing style that emerge within sub-communities are specific to those sub-communities (cf. Healey, 1997). The experiment takes place in two phases. In the first ‘convergence’ phase experimental sub-communities consisting of sub-groups of six people are formed. Subjects themselves are unaware of this sub-group manipulation, from their perspective the experiment consists of a series of rounds of Music Drawing with a different partner each time. However, during the convergence phase, the composition of pairs is controlled so that they are always made up of individuals from within the same sub-

group. This phase continues for four rounds thus allowing for a history of interactions to build up within each sub-group. On each round subjects perform the interactive version of the Music Drawing Task for 12 trials. Interactions in this first phase follow broadly the same pattern as simple interactive versions of the task. Subjects use Abstract drawings 63.7% of the time, Figurative 16.6% of the time (Composite drawings, a mixture of Abstract and Figurative elements, and trials with no drawing make up the remainder).

The second, experimental, phase occurs in the fifth round. In this round two conditions are compared; same-group pairs who are composed, as before, of subjects from within a single sub-group and cross-group pairs who are composed of subjects drawn from different subgroups.³ Same-group and cross-group pairs have equivalent task experience and expertise and, as noted, are unaware of any sub-group manipulation. Nonetheless they are reliably different in their use of the Drawing types. Multinomial regression analysis shows a reliable effect of the group manipulation on the distribution of Drawing types ($\chi^2_{(3)}=25.44, p=0.00$). The percentages are shown in Table 2.

Table 2: Use of Drawing Types in Pairs Drawn from the Same or Different Subgroups

Task Version	Drawing Type		
	Abstract	Figurative	Composite
Same-group	62.7%	11.1%	18.1%
Cross-group	41.3%	32.9%	15.8%

These results indicate that the co-ordination on particular drawing types that develops within Music Drawing Task sub-communities is a community-specific phenomenon. The conventions for graphical representation that develop within these experimental sub-communities do not readily transfer to interactions outside those sub-communities. Subjects in the cross-group interactions utilise a more mixed variety of drawing types which suggests that the representational strategies and conventions established within each group become unstable when people try to use them outside those groups. This finding parallels the results, reported in Healey (1997), which involve pairs communicating, verbally, about spatial locations. Following the same two phase experimental design, the results from this study also indicate that the types of spatial referring expressions established within particular sub-communities during the convergence phase were unstable in cross-group interactions.

Data from a community-based version of the concept communication task also indicate that communities of graphical communicators behave in a similar fashion to communities of linguistic communicators. The task requires a community group of 8 participants to communicate with each of the other 7 over an extended period of time. In the first round of the experiment the 8 work in 4 pairs with both

³The original design employed three experimental subgroups but for ease of exposition only two are reported here.

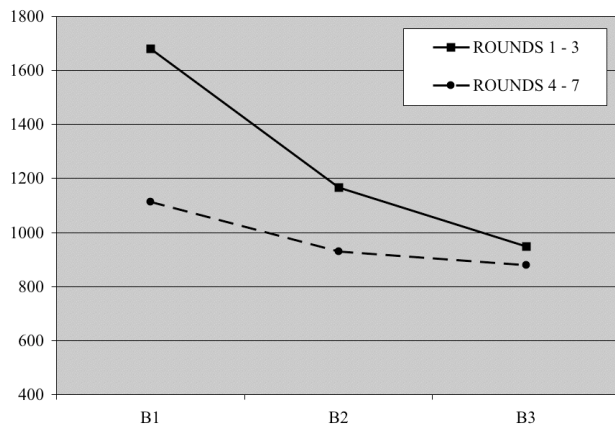


Figure 4: Reduction in Visual Complexity of Concept Drawings with Repetition in a Round (B1-B3 = Blocks of recurring items within a round)

participants drawing each concept 3 times over the course of the round. In the second round the 8 participants are re-paired and again draw the concepts 3 times. After each round they are re-paired again until every participant has encountered each of the others once and only once.

Figure 4 shows how the drawings become increasingly simple (according to the Pelli et al. measure) as the experiment proceeds. In the first 3 rounds this simplification process occurs across repetitions of the drawings (shown along the x axis of the figure). However, as the shared interaction within the community begins to develop (i.e., after round 4) the initial drawings in a round become as simple as the final drawings in the round. A similar pattern of results emerges for the communicators accuracy at identifying the concepts conveyed by their partners drawings. These findings are consistent with the idea that as a community becomes established through a common history of interaction so the drawings become conventionalised within the community: Drawings become simpler and more readily interpreted by the members of the community.

The implication of these results is that the processes which establish the conventions for producing and interpreting drawings and verbal descriptions operate in a manner that is directly tied to the character and pattern of interactions in which they were developed and used.

Discussion

Space constraints, and the ambition of providing an overview of a number of experimental results, dictate that much important detail has been elided from the descriptions of experiments and results provided above. Nonetheless, the results summarised above consistently point to the importance of interactional context for the interpretation of sketches.

Like referring expressions in conversation, the form and interpretation of drawings is systematically influenced by the character of the interaction in which they occur. Par-

ticipants in interactions show a strong tendency to match each others representational style and type. If items recur in an interaction, pairs also tend to develop increasingly abbreviated ways of representing them that are difficult for third parties to interpret. These patterns of change in the form of the sketches obtain independently of the particular concept or item being represented. In addition to these basic co-ordination processes of entrainment and abbreviation, there is also evidence that level of interaction available to participants affects the form of representations they use. In particular, the ability to localise, markup and re-draw elements of each others drawings appears to be important to the sustained use of more abstract representations. Lastly, this paper has presented evidence that interactions within sub-communities lead to the development of community-specific conventions for graphical interaction.

Together, these findings suggest that the ability to interpret sketches and drawings is substantially dependent on interactional context. It seems likely that this will prove to be an important issue for system design. Just as mechanisms of interaction are proving increasingly important to the development of effective natural language interfaces (cf. Trindi; Communicator), so sketch based human-machine interaction will need to address interactional context. In particular, a model of the mechanisms by which particular sketch interpretations are produced and sustained will be required.

As noted above, the parallels between graphical and verbal dialogues suggests the possibility that the findings reviewed above may reflect the operation of generic interactional mechanisms. For example, aspects of the experiments on graphical communication presented might be explained in terms of the collaborative model of grounding (Clark & Wilkes-Gibbs 1986; Clark 1996), input-output coordination (Garrod & Anderson 1987; Garrod & Doherty 1994), or conversational repair (Sacks, Schegloff, & Jefferson 1974; Schegloff 1992). The programmatic rationale for investigating tasks involving exclusively graphical communication is precisely that it has this potential to highlight basic mechanisms of communicative co-ordination that operate across modalities.

References

- Anderson, A., and Garrod, S. 1987. The dynamics of referential meaning in spontaneous dialogue. In Reilly, R. G., ed., *Communication Failure in Dialogue and Discourse*. Amsterdam: Elsevier. 161–183.
- Brannigan, H.; Pickering, M.; and Cleland, A. 2000. Syntactic co-ordination in dialogue. *Cognition* 75(B):13–25.
- Clark, H. H., and Wilkes-Gibbs, D. 1986. Referring as a collaborative process. *Cognition* 22:1–39.
- Clark, H. H. 1996. *Using Language*. Cambridge: Cambridge University Press.
- Clark, H. H. 1998. Communal lexicons. In Malmkjoer, K., and Williams, J., eds., *Context in language learning and language understanding*. Cambridge: CUP, 3rd edition. 63–87.

- Communicator: The DARPA Communicator Program: <http://fofoca.mitre.org/>.
- Engle, R. 1998. Not channels but composite signals: Speech, gesture, diagrams and object demonstrations are integrated in multimodal explanations. In Gernsbacher, M., and Derry, S., eds., *Proceedings of the 20th Annual Conference of the Cognitive Science Society*, 321–326.
- Garrod, S. C., and Anderson, A. 1987. Saying what you mean in dialogue: A study in conceptual and semantic coordination. *Cognition* 27:181–218.
- Garrod, S. C., and Doherty, G. 1994. Conversation, coordination and convention: an empirical investigation of how groups establish linguistic conventions. *Cognition* 53:181–215.
- Healey, P., and Peters, C. 2001. Notes on turn-taking and topic in drawing-in-interaction. Paper presented at the *The 4th International Workshop on Gesture and Sign Language based Human-Computer Interaction* City University, London.
- Healey, P.; Swoboda, N.; Umata, I.; and Katagiri, Y. 2001. Representational form and communicative use. In Moore, J., and Stenning, K., eds., *Proceedings of the 23rd Annual Conference of the Cognitive Science Society*, 411–416.
- Healey, P.; Swoboda, N.; Umata, I.; and Katagiri, Y. forthcoming. Graphical representation in graphical dialogue. Paper to appear in a special issue of the *International Journal of Human Computer Studies* on Interactive Graphical Communication.
- Healey, P.; Swoboda, N.; and King, J. 2002. A tool for performing and analysing experiments on graphical communication. Paper in preparation.
- Healey, P. 1997. Expertise or expertise?: The emergence of task-oriented sub-languages. In Shafto, M., and Langley, P., eds., *Proceedings of the 19th Annual Conference of the Cognitive Science Society*, 301–306.
- Hymes, D. 1972. Models of the interaction of language and social life. In Gumperz, J., and Hymes, D., eds., *Directions in Sociolinguistics*. New York: Holt, Rinehart and Wilson. 35–71.
- Krauss, R., and Fussell, S. R. 1996. Social psychological models of interpersonal communication. In Higgins, E.T. and Kruglanski, A., ed., *Social Psychology: Handbook of Basic Principles*. London: Guildford Press. 655–701.
- Neilsen, I., and Lee, J. 1994. Conversations with graphics: implications for the design of natural language/graphics interfaces. *International Journal of Human-Computer Studies* 40:509–541.
- Pelli, D. G.; Burns, C. W.; Farell, B.; and Moore, D. C. in press. Identifying letters. *Vision Research*.
- Sacks, H.; Schegloff, E.; and Jefferson, G. 1974. A simplest systematics for the organisation of turn-taking for conversation. *Language* 50:696–735.
- Scalife, M., and Rogers, Y. 1996. External cognition: How do graphical representations work? *International Journal of Human-Computer Studies* 45(2):185–213.
- Schegloff, E. A. 1992. Repair after the next turn: The last structurally provided defense of intersubjectivity in conversation. *American Journal of Sociology* 97(5):1295–1345.
- Schober, M. F., and Clark, H. H. 1989. Understanding by addressees and overhearers. *Cognitive Psychology* 21:211–232.
- van Sommers, P. 1984. *Drawing and Cognition: Descriptive and Experimental Studies of Graphic Production Processes*. Cambridge: Cambridge University Press.
- Trindi (Task Oriented Instructional Dialogue): <http://www.ling.gu.se/projekt/trindi/>
- Wilkes-Gibbs, D., and Clark, H. H. 1992. Coordinating beliefs in conversation. *Journal of Memory and Language* 31:183–194.