How People Produce Understandable Multi-Modal Explanations

Randi A. Engle (RANDI@CSLI.STANFORD.EDU)
Cognition and Learning Lab
Stanford University
Stanford, CA 94305-3084 USA

Introduction

The overall goal of this research is the development of a theory of communication that does not unnecessarily restrict itself to communication only via linguistic symbol systems (i.e., languages like ASL, English, or Japanese). I wish to contribute to the effort to explain how resources like diagrams, gestures, computer animations, and other non-linguistic signaling resources are used in coordination with language to achieve various communicative goals.

The communicative domain I have been focusing on are instructional explanations of mechanical devices. In part, because of the importance of spatial relationships between mechanical components in understanding how they operate, explainers frequently used iconic hand gestures to represent parts of the device or its motions, drew diagrams or gestured over complete ones when paper and pen were available, and operated and gestured over a sample device when it was made available. Besides contributing to the theory of communication, this work could have practical importance in the design of multi-media instruction and computer systems.

Data Collection

We first instructed six undergraduates in how such locks work and then videotaped each of them explaining the lock's mechanisms to three other undergraduate learners. For one explanation, a sample lock was available; for a second, pen and paper was available; for a third, no extra material resources were made available. Explainers and learners were told to take as long as they needed until they both felt that the learner understood how the lock worked well enough to be able to explain it to another person and thus so the learner would do well on a test to measure that skill.

Analysis of Explanations

All 18 explanations were fully transcribed. Explainers' and learners' speech was coded for pauses, hesitations, false starts, and intonational contours. All hand, head, and body, movements as well as manipulations of pen, paper or lock were described. The timing of these physical behaviors with respect to speech was noted to the nearest syllable.

One episode within each explanation—a conversational contribution to common ground (Clark & Schaefer, 1989) on a single topic—was selected for more intensive analysis. Half of the episodes focused on topics learners appeared to have understood on the basis of their re-explanations to explainers while the other episodes learners misunderstood in some way. Independent raters constructed situation models to represent what had been publicly communicated about the lock up to the point the episode began (the premodel) and by the end of the episode (the postmodel). For each transcribed behavior, raters also coded whether or not it played a communicative role in the explanation; and if communicative, whether it was hard to interpret for any reason.

Brief Overview of the Findings

1. Speech, gesture, diagram-drawing, manipulations of the lock are not treated by explainers, learners and raters as separate channels of communication that work separately but as part of larger units of communication Herb Clark and I are calling composite signals (Engle & Clark, 1995; Clark, 1996). Neither speech nor the non-linguistic signals are fully interpretable without each other.
2. In particular, explainers signal what signals are supposed to be part of the same composite signal by precisely synchronizing them to co-occur during the same time span. Similarly explainers use spatial contiguity between gestures, diagrams, the sample lock and other objects in common ground to put them into the same composite signal. If synchrony and contiguity are not sufficient, explainers can explicitly specify what is in the composite.
3. Each individual element of a composite is given a head start on its interpretation through the identification of what signaling method—conventional, indexical or iconic (Peirce, 1940)—is being used. If a conventional symbol is being used, interpretation involves accessing stored interpretations for that conventional symbol. If an index (like pointing) is being used, interpretation involves directing attention towards something in physical common ground. If an iconic symbol is being used, interpretation involves recognizing what the icon perceptually resembles.
4. Once the composite signal and the component signaling methods have been specified, this information can be used in the interpretation of each communicative element of the composite. Elements of the same composite are assumed to present consistent and complementary information about the same overall topic. This assumption allows interpretations to focus on the most appropriate conventional meanings, aspects of objects in physical common ground, and perceptual resemblances.
5. Finally, all this is done in order to allow the learner to build a workable model of the lock's operations from the explanation. Explainers build on common knowledge about how to use locks, produce composite signals tuned to the part of the explanation they are working on (abstract, spatial layout, causal chain), and revise their presentations until they are satisfied they make sense and the learner acknowledges understanding.