Researching Interactive Communication Behavior

A Sourcebook of Methods and Measures

C. Arthur VanLear
University of Connecticut

Daniel J. Canary
University of Utah
MICROANALYSIS OF FACE-TO-FACE DIALOGUE

An Inductive Approach

JANET BAVELAS, JENNIFER GERWING, SARA HEALING, AND CHRISTINE TOMORI

Microanalysis of face-to-face dialogue (MFD) began in the 1980s with a project on the curious phenomenon of motor mimicry, in which an observer responds in a way that would be appropriate to the situation of the person he or she is observing; for example, wincing when someone else is injured. Since Adam Smith (1759/2002, p. 12) described an example of motor mimicry, it had remained “a riddle in social psychology” (Allport, 1968, p. 30). Most theorists treated motor mimicry as an involuntary empathic reflex. However, in the course of 17 pilot studies aimed at capturing motor mimicry on video (Bavelas, Black, Lemery, MacInnis, & Mullett, 1986), we began to notice that it was more likely to happen when the other person would see it, which would not be true for a reflex.

Finally, on our 18th try (Bavelas, Black, Lemery, & Mullett, 1986), we discovered that motor mimicry was both extremely quick and highly social. This experiment was a carefully controlled 4-second sequence in which the participant was a passive observer. During the first 2 seconds, two experimenters were setting up equipment when one of them apparently dropped a heavy monitor on his obviously bandaged finger, then reacted by gasping and cringing in pain. His actions in the next 2 seconds created two experimental conditions: The experimenter either turned to face the participant, glanced at her, then made full eye contact (the experimental condition), or he turned away from her, toward the other experimenter, and interacted with him (the control condition). In the eye-contact condition, the participants started to display motor mimicry within an average of 1.04 seconds, and they continued with a kaleidoscope of mimetic expressions for a few more seconds; see Example 1 (in Figure 9.1). Without eye contact, participants either did not react at all or started a mimetic reaction and quickly stopped.

We began to appreciate that this rapid and precise sensitivity to the other person might be true of most communication in face-to-face dialogue (as our predecessors in the Natural History of an Interview project had proposed; Leeds-Hurwitz, 1987). However, this experiment was obviously not a dialogue. Because our ultimate interest was dialogue itself, we began to develop a method for the second-by-second analysis of face-to-face dialogues, for example, with speakers telling their own close-call stories rather than an experimenter enacting an injury. Example 2 (in Figure 9.2) illustrates an addressee’s motor mimicry while the speaker was telling his close-call story about an accident in the wilderness (see Bavelas, Coates, & Johnson, 2000).

What Is Microanalysis of Face-to-Face Dialogue?

The method that has evolved from that initial discovery is microanalysis of face-to-face dialogue, the detailed and replicable examination of any aspect of
observable communicative behavior as it occurs, moment by moment, in a face-to-face dialogue. Unlike many other methods for studying interaction behavior, MFD is not a coding procedure devoted to a particular variable or set of variables. Instead, it is an open-ended, inductive method for identifying and analyzing any behavioral, observable phenomenon within its theoretical scope. In this respect, MFD is similar to other meta-methods, such as conversation analysis, interactional sociolinguistics, action-implicative discourse analysis, critical discourse analysis, and discursive psychology (see Tracy, 1995, for a comparison of the other five approaches). Each of these meta-methods offers a theoretical and methodological framework for capturing any phenomenon a researcher finds interesting within its broad framework of assumptions.

To illustrate the wide applicability of MFD, Appendix 9.A describes the dozens of different analyses members of our research group developed for 24 projects on topics as diverse as conversational (co-speech) hand and facial gestures, ironic humor, dyadic decision making, generic versus specific listener responses, gaze patterns, positive versus negative topical content in psychotherapy sessions, collaboration in face-to-face versus computer-mediated communication, figurative language, early responsiveness of an infant with autism, patient-centered information in oncology consultations, formulations in psychotherapy sessions, and direct quotations. This chapter describes the theoretical and methodological choices (summarized in Box 9.1) that define and have shaped MFD.
Box 9.1
Characteristics of Microanalysis of Face-to-Face Dialogue

**Theoretical**

Discovering phenomena inductively, through observing the data
  vs. taking topics from the literature

Integrating spoken language with conversational hand gestures, facial gestures, and gaze
  vs. separating verbal and nonverbal communication

Collaborative model—dialogue as joint action; focus on the interactional function of acts
  vs. studying isolated acts of an individual or alternating monologues

Observable, functional perspective on interaction
  vs. inferences or attributions about cognition, motivation, emotion, etc.

**Requirements**

Spontaneous\(^a\) face-to-face dialogue
  vs. interaction with a confederate or the experimenter

Video recording (preferably with all interlocutors on camera)
  vs. audio recording or transcript

Analysis linked to video (preferably annotation of digitized video)
  vs. annotation on a transcript or in a separate document

Analysis of moment-by-moment details, often in fractions of seconds
  vs. paraphrasing, summarizing, or rating

Analysis takes into account the accumulating conversational context
  vs. decontextualized utterances

**Options**

Location can be in a lab experiment or field observation
  vs. either exclusively experimental or exclusively “naturalistic”

Research design can be experimental, quasi-experimental, or no experimental design
  vs. either limited to experiments or excluding experiments

(Continued)
An Inductive Approach

In research, to *induce* is “to infer from particulars” rather than using a priori or preexisting principles. For example, although Sherlock Holmes is usually described as “deductive,” he consistently formed his hypotheses inductively by observing the particulars closely and making inferences from them rather than by relying on what he had been told about the situation.

Being inductive does not mean having a blank mind; one always starts with interests and curiosities within an explicit or implicit theoretical framework. Being inductive means building theory up from the particulars of observed data. Inductive discovery is valued in many scientific fields, from Francis Bacon (1620–1911) to Nobel Prize winners such as Sir Peter Medawar (1979), Barbara McClintock (in Keller, 1983), and Rita Levi-Montalcini (1988).

An inductive approach is rarely reflected in the text of published articles because their function is to report results efficiently within the context of existing theory and literature rather than tracing the chronological development of projects with inductive origins. Although some methods texts and research courses mention inductive methods, a great deal of what students learn to do leads them away from beginning inductively. Box 9.2 summarizes the guidelines from an essay on inductive research (Bavelas, 1987), with suggestions on how to nurture a new idea as well as how to avoid losing it.

The inductive phase precedes and supports—but does not replace—the later, deductive phase of critical assessment and hypothesis testing. When we have used MFD for formal hypothesis testing, our hypotheses often came from repeated viewing and analysis of video-recorded dialogues. Appendix 9.A notes the studies that started inductively, that is, where the phenomenon itself (or the approach to it) came from observing data rather than reading the literature. Although the relevant established literature can provide some necessary background, it is close observation that directs the researcher’s attention to patterns of behaviors that could be fruitful for systematic analysis. Decisions about what could be of particular interest (and why) arise in the relationship between curiosity and direct observation.

Starting inductively has advantages. First, face-to-face dialogue is so intricate and complex that innumerable findings still await discovery. As an alternative to seeking gaps in the literature that might need filling, observing dialogues directly can expand the literature and even take it in new directions. Second, building a theory from data creates a secure empirical base and acts as a restraint on overgeneralization. There is a risk that “grand theories” may achieve their scope more by the encompassing language of the theory than by the data that support it. Third, exploring hunches to find new phenomena is scary but also exciting. Testing these hunches by watching and listening to a dialogue can be far more motivating than using a “top-down”
Box 9.2
Some Guidelines for the Inductive Phase of Research

First, you notice something that intrigues you . . .

1. Don't dismiss it (e.g., concluding it's just "chance" or misperception).
2. Don't go find a category to put it in (e.g., naming it with a technical term from the literature or an expert).
3. Don't belittle it (e.g., seeing it too narrowly or literally).
4. Don't be practical or critical (e.g., worrying about all of the problems that lie ahead).
5. Don't panic (e.g., rushing into a study too soon).

All of those will kill your intuition for sure. Instead you should proceed slowly:

1. Do get more examples: Collect observed examples; no verbal labels yet. Call them instances of "X."
2. Do start developing a schema or class. Consider what these examples might have in common.
3. Do articulate by analogues. Think in metaphors ("It is as if . . . "), not technical language.
4. Do unearth the model. Be an archaeologist, uncovering what is there.
5. Finally, do start pilot work. Arrange to see more of X.

Now it is safe to start putting X into words, preferably a full phrase, describing its function, not just a single term.

Source: Adapted from Bavelas (1987), where these steps are described more fully.

approach that starts with someone else's theory. Finally, the focus on the micro-details of dialogue pays off—one sees a new phenomenon directly and can show it to other researchers, who can then see it as well. The "new" phenomenon was always there; we just had to notice it.

Theoretical Framework

All observation is selective, and the decisions that shape any analysis inevitably reflect the analyst's theoretical assumptions (e.g., Ochs, 1979). MFD rests on two theoretical assumptions about face-to-face dialogue (described in more detail below) that have been the focus of our long-term program of research. First, because interlocutors in face-to-face dialogue can see and hear each other, they use both visible and audible communicative resources, which are tightly integrated with each other. Second, because interlocutors are engaged in social interaction, their actions must be understood as coordinated and mutually influential.

Integrated Audible and Visible Acts

One major theoretical assumption of MFD is that face-to-face dialogue consists of highly integrated combinations of audible communicative acts (words and prosody) with a defined set of visible co-speech acts. Birdwhistell (1970) proposed that communication consisted of "a structural system of significant symbols (from all of the sensorily based modalities) which permit ordered human interaction" (p. 95), and subsequent scholars have used terms such as ensembles (Kendon, 2004), composites (Clark, 1996), and integrated messages (Bavelas & Chovil, 2000, 2006) to describe these
multimodal acts. This position contrasts with some other approaches that, implicitly or explicitly, treat all nonverbal behaviors as either redundant with speech or as conveying an entirely different kind of information.

Bavelas and Chovil (2006) specified conversational hand gestures, conversational (nonemotional) facial gestures, and mutual gaze as the visible co-speech acts that contribute to meaning in face-to-face dialogue because they are tightly synchronized with speech in both timing and meaning. Conversational hand gestures are rapid and precise enough to be timed within fractions of seconds of the words they are related to (e.g., Kendon, 1980; McNeill, 1992; Nobe, 2000) as well as adapting to disfluencies (Seyfeddinipur, 2006). Chovil (1989, 1991/1992) found that conversational facial displays are timed to a single word or even syllable. The meaning of a hand or facial gesture is also synchronized with the words it accompanies, conveying information that supplements or even complements (i.e., is not redundant with) these words. Hand and facial gestures can contribute both semantic content and pragmatic information about the ongoing interaction. Gestures are “visible utterances” (Kendon, 2004), and the majority of facial actions in dialogues are not about emotion (Fridlund, Ekman, & Oster, 1987, p. 160). (See Bavelas, Gerwing, & Healing, 2014b, for a review of research on conversational hand and facial gestures.) Patterns of mutual gaze are also precisely timed and tend to serve interactive functions, especially marking speaking and listening roles (Argyle & Cook, 1976; Duncan & Fiske, 1977; Kendon, 1967).

MFD includes these co-occurring audible and visible communicative acts as an integrated whole and interprets them in relationship to their immediate conversational context. Three examples will illustrate conversational hand gestures, facial gestures, and gaze, as well as providing an introduction to the micro level of our analysis, which is often frame by frame.

Example 3: A Conversational Hand Gesture

Two students were being video-recorded as part of a psychology study. They began by getting acquainted, and one of them explained how he came to be in this study:

**Speaker:** “I, uh, I’m taking a couple of, uh, **Psych courses—Drugs and Behavior**—so that’s how they got my name.”

**Addressee:** “So that’s how they got yours, huh.”

The underlining indicates where the speaker gestured a writing motion, as if filling out the volunteer form handed out in all psychology courses (in his case, Drugs and Behavior), which led to his being called for this study. Figure 9.3 shows frame shots of this sequence. The speaker’s hands were clasped loosely in his lap, then in the 1.37 seconds it took to say “so that’s how they got my name,” he raised his right hand, mimed a writing motion, and returned his hand to its original position. The writing motion itself was 0.36 seconds and was

![Figure 9.3 Complementary function of a hand gesture. Example 3, the speaker uses a nonredundant hand gesture. The speaker and addressee were sitting directly across from each other. Frame 1 shows the split-screen configuration with both of them; Frames 2 and 3 show only the speaker’s side. His writing motion in Frame 2 conveyed essential information that was not in his words, that is, that he had filled out a volunteer form.](image-url)
timed precisely with the words “how they got.” Notice that his words did not mention filling out the form, so without the gesture, “that’s how” would be unclear. This is an example of a nonredundant gesture, which provides relevant information that the accompanying speech does not convey. In face-to-face dialogues (vs. on the phone), nonredundant gestures are significantly more frequent, and speakers convey less of the essential information in their speech (e.g., Gerwing & Allison, 2011).

Example 4: Two Functions of Conversational Facial Gestures

The speaker had watched scenes from the movie Shrek 2 and was now recounting them to the addressee. At one point, she described a scene in which Shrek was talking to Donkey about Puss in Boots:

“And Shrek’s like, ‘How can we not take him, he’s so—look at him, he’s so cute, Ohhh. How many cats can wear boots?’ And then [pause, 0.77 seconds], I guess they take him [pause, 0.67 seconds] and that’s the end.”

During the first three underlined phrases, the speaker was not only quoting Shrek but was also portraying him with her head and face by looking down (as Shrek had looked down at the tiny cat) and pursing her lips as she spoke in a singsong voice. The second frame in Figure 9.4 shows one of these portrayals. Then, during the two underlined pauses, her facial gestures served a different function. Both of these were facial shrugs (Chovil, 1989, 1991/1992): She looked away from the addressee, flashed her eyebrows, and quickly pulled one corner of her closed mouth down and then back to level (e.g., the third frame in Figure 9.4). Like a shoulder shrug, a facial shrug indicates uncertainty or dismissal. Listening to an audio recording would leave the two pauses open to interpretation, whereas the video shows that she was indicating to the addressee that she had nothing more to say. (See Chovil, 2005, for further methodological details.)

Example 5: A Conversational Function of Gaze

The speaker was telling a close-call story in which the heat from a reading lamp on her headboard set her pillow on fire as she slept. She had already explained where the lamp was and how hot it was. Now she began to foreshadow the close call:

I guess it was on for—I don’t know how long it was on for.

Figure 9.5 shows the gaze pattern for this sequence. Notice that in Frame 1, the speaker is looking down, and the addressee is holding her previous expression (two fingers near her mouth). In Frame 2, as she said the words underlined above, the speaker looked up and came into full eye contact with the addressee, who immediately raised her eyebrows noticeably, creating a wide-eyed and somewhat alarmed facial gesture, which she held for 1 second, still saying nothing. Bavelas, Coates, and Johnson (2002) showed that these brief moments of mutual gaze during the speaker’s turn have a very high probability of eliciting a new response from the addressee.

The above examples illustrate that a theory of integrated audible and visible co-speech acts has clear

---

**Figure 9.4** Two complementary functions of facial gestures. Example 4, two nonredundant facial gestures. Both conveyed information that was not in the speaker’s words. (The addressee, who was sitting directly across from the speaker, is inset in the circle in the upper right corner.) The speaker was describing a scene from the movie *Shrek* 2, and in Frame 2, she enacted Shrek talking about the cat while looking down at him, with pursed lips and a doting expression. Immediately after, in Frame 3, she paused, made a quick facial shrug, and provided no additional details.
methodological implications: Analyzing an audio recording or a transcript of a face-to-face dialogue will exclude meaningful information that is likely to affect the interpretation of the words themselves.

Dialogue as Joint Action

Our other major theoretical assumption owes a great deal to the collaborative theory of Clark and his colleagues (e.g., Clark, 1992, 1996). This theory casts dialogue as coordinated joint action between speaker and addressee rather than as autonomous actions by individuals. Experiments by Clark and Wilkes-Gibbs (1986), Isaacs and Clark (1987), Schober and Clark (1989), and Bavelas et al. (2000, 2002) have shown that a collaborative model predicts what happens in dialogue better than an autonomous model does.

The above dialogue examples all illustrated this moment-by-moment coordination:

- In Example 2 (Figure 9.2), the addressee’s wince and “Ooooh” were not his individual emotional reactions; they were perfectly timed to fit a particular moment in the speaker’s narrative.
- In Example 3, the speaker’s gesture drew on their shared knowledge. Notice that the addressee collaborated by confirming that he had indeed understood, using the speaker’s exact words, including the demonstrative pronoun referring to the gesture (“that’s how they got”).
- In Example 4, the facial shrug was important information for the addressee. It accounted for the speaker’s momentary pause and also qualified what she said next as inexact but good enough.
- In Example 5, the speaker’s gaze elicited the addressee’s response (looking alarmed), which in turn showed the speaker that the addressee was following her story closely without interrupting it.

An Observable, Functional Perspective on Interaction

Studying dialogue as collaborative interaction limits both theory and data to the analysis of each act’s function in the dialogue, that is, the moment-by-moment effect of what the interlocutors display to each other, as the dialogue unfolds. A particular project may focus only on the speaker or the addressee, but our collaborative approach directs the focus outward to what the action is doing in the dialogue rather than inward to what it might mean about that individual. MFD explicitly avoids inferences about unobservable mental processes (e.g., an individual’s cognition, intention, motivation, emotion, attitude, ability, or personality) and instead focuses on how an act responds to what happened before or leads to what happens next.

We propose that careful documentation of what is observable in dialogue should precede speculation about possible mental or neurophysiological explanations.
Chapter 9 Microanalysis of Face-to-Face Dialogue  137

Data Requirements: What Is Essential and What Is Not

As a practical matter, it is possible to embark on the study of observable behaviors only with suitable data and technology. Fortunately, the minimum requirements (outlined in this section) are neither onerous nor expensive.

Spontaneous Face-to-Face Dialogue

Because MFD was developed for the analysis of face-to-face dialogues, this method requires spontaneous dialogues in which participants can see and interact with each other. Most obviously, the data must be social behavior in a dialogue, not the increasingly common use of self-reports or responses to stimuli on a computer (Baumeister, Vohs, & Funder, 2007; Patterson, 2008; Patterson, Giles, & Teske, 2011). Other settings, such as telephone or mediated dialogues, are not inherently less interesting or informative, but one cannot assume without evidence that they will generalize to face-to-face dialogues. Indeed, their value may be the contrast they provide to a face-to-face dialogue (e.g., Bavelas, Gerwing, Sutton, & Prevost, 2008; Bavelas, Gerwing, & Healing, 2014a; Phillips, 2007).

A defining feature of face-to-face dialogue is the participants’ reciprocal ability to act and react spontaneously. In Clark’s (1996, p. 10) terms, a true dialogue requires extemporaneity, self-determination, and self-expression. Even though dialogues in experiments typically have a particular focus or task that directs the topic of their conversation (e.g., getting acquainted, telling a story, giving directions), ensuring spontaneous interaction within that task is still essential. Social psychology experiments frequently seek to trade off spontaneity for experimental control by using a confederate or experimenter as one of the participants. However, Kuhl and Brennan (2013) have raised serious questions about actually achieving experimental control with confederates. There is also evidence that the results with confederates or experimenters may differ from the results with free dialogues (Bavelas & Healing, 2013; Brown-Schmidt & Tanenhaus, 2008; Lockridge & Brennan, 2002). Ethical considerations, such as in medical interactions, often mandate analog or simulated dialogues, but these also raise questions about control and generalizability to free dialogues (e.g., de la Croix & Skelton, 2009; Clever et al., 2011). Fortunately, more and more lab experiments on language and communication are using free dialogues, which demonstrates that shifting the unit of analysis from the individual to the dyad does not compromise experimental control.

Video Recording of Both (or All) Participants

In face-to-face dialogues, the participants can see and hear each other (Clark, 1996, p. 9), and video recording is the only method that preserves the participants’ integrated audible and visible communicative acts. Working with an audio recording or a transcript of a face-to-face dialogue implicitly assumes that visible behaviors are either redundant with speech, are irrelevant, or are a separate “channel” conveying information that is unrelated to the co-occurring speech (e.g., emotional or relationship information). Yet some visible acts, such as nodding, are directly equivalent to verbal alternatives such as “yeah,” and omitting them creates an apparently less responsive addressee. Similarly, as illustrated in Examples 3 to 5, leaving visible co-speech acts unrecorded will often lead to erroneous or ambiguous interpretations of the speech itself. Even some basic assumptions may no longer apply. For example, observing how the addressees in Examples 2 and 5 provided overlapping but noninterruptive facial reactions to their speaker makes it hard to maintain an individual’s speaking turn as the basic unit of dialogue.

A less obvious reason for video recordings is that these capture an often overlooked feature of copresence, namely, the physical and social context that the participants share, including how they use objects and refer to what is mutually visible. When listening to an audio recording or reading a transcript, it is easy to forget whether the participants were in a medical office, a living room, a bar, or a psychology experiment, much less how they used that environment as part of their dialogue.

Studying face-to-face dialogue requires that both (or all) participants be on video at all times. In the lab or a setting in which seating can be assigned, a precisely planned side view with a single camera will capture two participants. Two or more cameras and a special effects generator can create a variety of split-screen configurations, but a well-placed mirror can also create a split-screen image with one camera (Bavelas, Black, Lemery, MacInnis, et al., 1986, pp. 107–108). Luff and Heath (2012) provided an excellent discussion of the possibilities for field studies. Getting both participants on screen is more than a technical issue; it also reveals how theory can shape method even at the data collection stage. For example, switching the camera back and forth to record only the person who is speaking at the time imposes a theory that treats dialogue as alternating monologues.
Analysis of such recordings preserves the autonomous model, as there is no record of the listener’s visible behaviors, let alone their effect on the speaker. For example, in a classic study, Goodwin (1981) challenged those who interpreted speakers’ frequent sentence fragments in spontaneous dialogue as evidence of “the apparent disorderliness of natural speech.” By video-recording both participants, Goodwin showed that speakers broke off mid-sentence when their addressees were not looking at them and then restarted when they had established mutual gaze—an orderly collaborative pattern. A camera view that focuses only on the speaker makes it difficult if not impossible to document such collaboration.

Any Setting, Any Design

Assuming that the aforementioned requirements are feasible, any setting and any research design will work. Face-to-face dialogue is the pervasive form of language use (e.g., Goodwin, 1981, p. 12; Goodwin, 1990, p. 1; Levinson, 1983, p. 284; van Dijk, 1985, p. 2) and is remarkably adaptive to the particular contexts in which it occurs. Given this diversity of contexts, no single setting has a claim to being the “real world,” and none can be dismissed as “artificial.” We do primarily lab experiments but have also studied oncology consultations (Healing, 2013), hospital interactions (Gerwing & Dalby, 2014), psychotherapy sessions (Korman, Bavelas, & De Jong, 2013; Phillips, 1998, 1999; Tomori, 2004; Tomori & Bavelas, 2007; Smock Jordan, Froerer, & Bavelas, 2013), and home videos of infants (Gerwing, 2008).

As for research design (e.g., a randomized controlled experiment, a quasi-experiment, a field study, or pure observation), the choice will primarily affect the interpretation of results, that is, internal versus external validity. It is probably more accurate and fruitful to combine curiosity about others’ data in different settings with more modest generalizations about our own.

Using Video Annotation
Rather Than a Transcript

For many years, our research group used a variety of low-tech methods, mostly looking at analogue video and recording our decisions on paper. The current variety of video annotation software systems offer distinct advantages, but researchers should approach these systems with specific questions: First, if the cost of the software and support is covered by a site license at the researcher’s institution, what happens if the researcher moves elsewhere? Or if the institution decides to let the license expire? The researcher’s previous or ongoing data analysis using this system could be lost or made unavailable. Second, does the software accommodate a novel analysis system, or does it impose fixed coding criteria? Does it allow the researcher to work inductively, being responsive to the data and adapting to what it is revealing? Annotation software should support both the inductive phase and the final, formal, or deductive phase.

Our research group has settled on ELAN software for annotating digitized video (http://tla.mpi.nl/tools/tla-tools/elan/; Wittenburg, Brugman, Russel, Klassman, & Sloetjes, 2006). ELAN is open-source, free software developed and maintained at the Max Planck Institute for Psycholinguistics in the Netherlands. ELAN is a highly flexible and intuitive tool for the analysis of most digitized formats and has several valuable functions. As shown in the screenshot in Figure 9.6, the analyst can play and move around the video at a variety of speeds, including fractions of seconds. Even more important, dragging the mouse along the time line (creating the blue column) selects any section of dialogue for repeated or frame-by-frame viewing, as well as preserving it. Selection is the core of MFD, because it reveals details of sequences that are often not obvious in a first or real-time viewing. ELAN is not only the microscope that takes the analyst into the world of these details, it is a method for preserving them.

Figure 9.6 shows the tiers (rows) below the video where the researcher can annotate multiple, even overlapping behaviors. For example, the speaker’s actions can be noted on one tier, and overlapping addresstee contributions can be annotated on another tier. The researcher can add as many tiers as needed and label them as he or she wishes; there are no imposed units, features, or variables. Figure 9.6 includes a simple illustration, showing how ELAN captured the information needed for Example 3 (i.e., the exact timing of the speaker’s gesture to his words). First, the analyst had to locate where his entire hand movement occurred, from the moment he raised his hand to when he put it back on his lap. Using ELAN with the sound off, the analyst located those two points by moving backward and forward repeatedly, then selected that segment and saved it on the top tier, which she named “full gesture phase.” She could then listen to this selection repeatedly to be able to record the precise speech accompanying the movement into this tier. More repeated viewing located exactly where the meaningful part of the gesture began and ended, as well as the exact words it accompanied. This smaller selection was saved on the second tier, using the technical term gesture stroke. These two tiers could later be used to record all of the gesture-word combinations in this dialogue as part of a fuller study of the precise synchronization of gestures to words.
Figure 9.6 Screenshot of ELAN software in use. The ELAN screen includes the digitized video, action buttons, and tiers for annotation. The action buttons directly below the video include play (or pause) and frame-by-frame forward or backward. The buttons to their right are for playing a selected excerpt. The tiers below illustrate the selection and annotation used to analyze Example 3. The upper right quadrant includes a choice ("Grid") to view all annotations on a given tier.

For MFD, the essential feature of ELAN is that it links the words, prosody, and visible acts just as they occurred in the dialogue. Relying on a separate written transcript of the video recording may lead to analyzing the transcript rather than the dialogue. Although it is quicker to read a transcript than to look at a video, one’s memory of the visible and prosodic features is fallible. Even notations of these features on the transcript are still only the transcriber’s version of the data. ELAN displays what the interlocutors actually did.

The Scale of Microanalysis

As illustrated in Examples 1 to 5, "micro" in our method means attending to the observable moment-by-moment details of dialogue, often at the level of fractions of seconds. The goal is to discover the precision with which the participants create and manage their dialogue. Therefore, the primary raw data for microanalysis are these details, not the analyst’s paraphrase, summary, or linguistic gloss. The latter are processed versions of what happened, usually at a theoretical or conceptual level.

Although the focus of MFD is the specific behavior or behavioral process that is of interest in a particular project, it is essential to situate each instance in a nested set of contexts. These include the setting of the dialogue in the broadest sense (e.g., a medical consultation, a psychology experiment, or home videos of infants), who the interlocutors are (e.g., physician and patient, strangers meeting for the first time, infant siblings), the purpose of their dialogue (e.g., diagnosing the patient’s illness, getting acquainted, playing with new toys), the accumulated dialogue so far (e.g., the topics covered, who has said or done what), and finally the micro-moment in which a particular instance happens. As the dialogue proceeds, the participants inevitably draw on and create shared understandings, which shape their subsequent actions and meanings. For example, when information is new to the dialogue, the participants tend to use full noun phrases and to articulate the information clearly. Later, when that information becomes “given,” participants often replace nouns with pronouns or other shorthand references and even speak less clearly (Fowler, 1988). In a larger sense, what either person says builds on the dialogue they have created together so far.

Overlapping Phases of an Inductive Microanalysis

Many researchers approach video-recorded dialogues with their own particular curiosity, like a detective seeking to uncover specific information.
The curiosity might be about how a particular kind of communicative act (e.g., hesitation, metaphor, disagreement) functions in dialogue or about how a particular interactive process unfolds during dialogue (e.g., cooperating on a task, constructing the solution to a problem). The path that starts with curiosity goes through definition and analysis, until it reaches its goal of satisfying the curiosity and explaining it to the wider world. Be aware that this path is rarely as straightforward as can be presented here; the phases will overlap and often involve backtracking. In this section we illustrate several stages on the path for Healing’s (2013) master’s thesis on oncology consultations.

On Being Inductive

In the study of face-to-face dialogue, being inductive means developing an analysis from the observable details of interlocutors’ communicative behaviors, rather than relying solely on previous work or existing theory. Being inductive does not mean avoiding any theoretical framework; ideally, the researcher thinks globally and acts locally. As noted at the outset, MFD includes a global framework that suits our interests; focusing on local details within this framework can lead to innovation. Being inductive does not mean abandoning one’s training in rigorous research methods for a carefree romp through the data. Nor does it mean passively waiting for inspiration to strike. It means adopting the equally demanding mental discipline of patience, of initially postponing those rigorous, hypercritical research methods that will be necessary later, in order to engage with the data intensively and freely at the beginning.

Healing (2013) explicated and documented (http://dspace.library.uvic.ca/bitstream/handle/1828/4835/Healing_Sara_MSc_2013.pdf?sequence=7&isAllowed=y) the inductive process that led to her analysis of the information patients contributed during their oncology consultations. As part of a larger project at a cancer agency, Healing had the opportunity to sit in on oncology consultations when the physician and patient agreed. She soon became interested in shared decision making. The oncologists contributed essential information relevant to treatment options, but Healing was curious about the patients’ unique contributions about their own situations (e.g., what a particular patient valued or feared). She proposed that a truly shared decision process should, at a minimum, include contributions from the patient that were relevant to the decision. Healing found that the literature on patient-centered communication focused on how health care professionals should talk to patients but not on what an individual patient can contribute to the dialogue. She was confident that she could find the answer inductively, through close examination of video-recorded consultations. After the necessary ethical and logistical steps, she was able to video-record eight full oncology consultations.

In these consultations, Healing’s (2013) raw data consisted of everything each patient said, so she began by creating an ELAN tier with an entry each time a patient said something: 1,585 utterances. Then she immersed herself in a subset of about 30% of her data until she began to be able to articulate patterns in what patients were contributing, mainly by making notes on the function of what the patient said. This was a time-consuming, trial-and-error process, but it brought the patient’s contributions out of the background and began to reveal their variety.

The first major distinction that emerged was between two different functions: Naturally, many utterances contributed biomedical information about the patients’ cancer, treatment, or other procedures (e.g., telling the physician what tests they were taking). Other utterances functioned to introduce patient-centered information, which she defined as information that contributed this particular patient’s unique position or point of view. Patient-centered information included, for example, factors affecting this individual’s choice of treatment:

- “We were kinda hoping, in September, to go on a cruise” [i.e., it would be better to have treatment concentrated in a shorter time].
- “So, yeah, um, and a, I like to, ah I don’t expect to be a rocket scientist so I want a brain that will function for everyday activities. I like to do photography and, a, I like being involved, you know, in art things and whatnot.”
- “Um, I pretty well came here with a totally open mind [about treatment options].”

Other topics were the patient’s limitations or resources:

- “I seem to ah, if, if ah drugs have a downside, or, or a counter-effect thing I seem to get them. The, the Statin drugs are really a problem for me.”
- “And since I’ve got so many buddies who’ve been through systems and everything, ah yeah, very good support system.”
Patients also talked about their hopes and fears:

- “If I make it into my eighties, kickin’ and screamin’ then that’s good.”
- “That’s probably the thing that scares me the most ah, is incontinence.”
- “[When the other doctor said] that the chances of survival are 80%, I said, ‘well gee, I’ll take the 80%’ [slight laugh].”
- “If, if I’m, if things are so bad—I’d just as soon take a potion and exit.”

Healing’s immersion in the data identified two more functions of client’s utterances: small talk, which contributed information not directly related to the illness or the treatment, and generic listening responses (e.g., “yeah” or “mm hmm” when these simply showed that the patient was listening and following what the physician was saying). However, like the biomedical information, these were not the focus of her research. Instead, she defined and identified these three functions in order to provide contrasts that sharpened the definition of her primary interest in patient-centered information.

Developing an Operational Definition

Once the researcher’s immersion in the data has led to an intuitive recognition of the phenomenon, it is time to start thinking about how to articulate it for other researchers, that is, to develop an operational definition. An operational definition connects the researcher’s broad, usually abstract definition of the phenomenon with specifics about how to recognize the phenomenon in a dialogue. Formally, an operational definition is a set of rules and procedures (ideally with many examples) that permit others to locate instances such as those the researcher has intuitively uncovered (e.g., Healing, 2013, pp. 27–38). A fully explicit operational definition serves to share knowledge with other researchers, to permit others to check on the researcher’s conclusions, and to encourage replication. As a bonus, the process of becoming very explicit inevitably leads to a deeper understanding of the topic itself.

Creating a logical sequence of decisions. A typical analysis requires more than one decision about the same section of data. Most often, the first task is simply to locate all instances of the phenomenon (e.g., all patient utterances). The second decision is to characterize each of these instances according to the operational definition (e.g., as small talk, a generic listener response, biomedical information, or patient-centered information). Locating and characterizing phenomena at the same time conflates decisions that should be separate and makes disagreements between analysts difficult to interpret. For example, if one of the analysts overlooked a patient utterance, but the other one analyzed it, they would inevitably (and spuriously) disagree about what its function was. The best procedure is to complete one level of decisions, assess agreement, and resolve any disagreements before proceeding to the next level.

Sometimes the analysis requires several contingent decisions; that is, one decision determines what the choices will be for the next decision. Constructing a formal decision tree that describes these branching decisions is a useful addition to an operational definition because it ensures that each analyst follows the same logical decision processes. Appendix 9.B includes decision trees from three different projects, including Healing (2013). In each example, the easiest decisions come first; these usually locate instances that clearly do or do not fit the operational definition. In Healing’s data, utterances that were small talk or generic listener responses were easy to recognize, so identifying them and removing them from further consideration left the analyst free to focus on the more difficult distinctions between biomedical and patient-centered information.

Interanalyst Agreement and Cross-Validation

Two criticisms or concerns often arise about inductive research. The first is that the findings could be idiosyncratic to this researcher and no one else could see them. The answer is high inter-analyst agreement using the operational definitions. On the basis of approximately 15% of the data that were not part of the inductive data set, Healing (2013) demonstrated 90% agreement (205 of 228 decisions) with an independent analyst. The second concern is that the finding might fit only the specific data set it came from and would not hold in other relevant data. The answer is cross-validation with new data. If the analysis fit only the specific data set it came from, then it would not apply to new data, leading to numerous cases that were “not analyzable” or “other.” Healing’s cross-validation included the reliability data set plus the remainder of the data (which she analyzed herself). Over 96% of these 1,135 utterances fit one of the four functions she defined in her analysis. Her analysis was ready for use in new research projects or for teaching physicians to recognize patient-centered information.
Concluding Thoughts

In this chapter we have described and illustrated microanalysis of face-to-face dialogue, which is an open-ended, inductive method for the detailed and replicable examination of any aspect of observable communicative behavior in its immediate context. We have been explicit about the theoretical framework of this approach as well as the data it fits and as explicit as possible about the process of doing research inductively.

Some nontechnical advice may be helpful as well. First, it is essential to recognize that the inductive phase is a distinct initial phase of a process that eventually leads to the more familiar, formal procedures (e.g., design and hypothesis testing). However, in the early phase, it is often difficult to put aside those formal criteria, simply out of fear. After all, methods textbooks and courses often overwhelm students with all of the possible errors and misinterpretations lying in wait for a researcher (e.g., alternative explanations, Type I error). If internalized, this kind of criticism is crippling during the early phases (e.g., Bavelas, 1987). A clear recognition that those formal issues will come later can allay the fear, leaving the researcher free to follow his or her intuition and passion in the early phases.

A second question arises because MFD (as some other similar approaches) is so time-consuming. More than one colleague has suggested that a "quick and dirty" analysis would be just as good. Curiously, no one suggests faster alternatives in other sciences, such as neuroscience, deep-space astronomy, and higher mathematics, in which painstaking and time-intensive analysis is the norm. Some phenomena are simply not visible except at a micro level. For example, Gerwing's (2008; https://dspace.library.uvic.ca/handle/1828/1226) data were home videos of infant triplets, one of whom was later diagnosed with autism. Several professionals had watched these videos in real time, and none had identified which infant would later be diagnosed. Gerwing created an intensive microanalysis of each infant's immediate response to any interactional overture by a parent (see Appendices 9.A and 9.B) and applied it twice with "blind" analysts. The analysis not only differentiated between the infant with autism and his typically developing siblings, it also documented when the differences from his siblings appeared and in which kinds of interactions. In our experience, the seconds and fractions of seconds of a face-to-face dialogue often contain interactional patterns and processes discoverable only with a "slow but clean" microanalysis.

References


## Appendix 9.A

**Studies by Our Research Group Using Microanalysis of Face-to-Face Dialogue**

<table>
<thead>
<tr>
<th>Year</th>
<th>Study</th>
<th>Inductive Origin?</th>
<th>Analysis</th>
<th>Interanalyst Agreement for Each Analysis</th>
</tr>
</thead>
</table>
| 1989; 1991/1992 | Chovil (doctoral dissertation; journal article) | Yes               | What kinds of conversational facial displays occur in dialogue? (This was the first systematic study of nonemotional facial actions.)  
   a. Whether a *conversational facial display* occurred  
   b. Deciding on the general function of the display (*syntactic, semantic, or nonlinguistic*)  
   c. Deciding on the specific function of the display (e.g., *question marker, portrayal, adaptor*) | a. 90%  
   b. 96%  
   c. 87% |
| 1989; 1991 | Chovil (doctoral dissertation; journal article) | Yes               | Does visibility or being in a dialogue affect the frequency of addressees’ facial motor mimicry?  
   a. Number of *motor mimicry* displays per listener | a.  $r = .94$ |
| 1991     | Coates (master’s thesis)            | Yes               | How do participants collaborate on irony in dialogue?  
   a. Identifying inversions (*instances of ironic humor*)  
   b. General meaning of an utterance: *mock vs. serious*  
   c. Specific meaning of an utterance (14 categories, including *serious excitement/mock excitement, serious insult/mock insult, serious/mock horror*, etc.)  
   d. Identifying the four phases of an inversion sequence (*implicit agreement, delivery of inversion, displaying understanding, and closure*) | a. 92%  
   b. 100%  
   c. 88%  
   d. 96% |
| 1992     | Bavelas, Chovil, Lawrie, & Wade (journal article) | Yes               | What is the nature of hand gestures that are not depicting topical content? (See decision tree in Appendix 9.B.)  
   Experiment 1.  
   Locating *topical vs. interactive gestures* | |

*(Continued)*
<table>
<thead>
<tr>
<th>Year</th>
<th>Study</th>
<th>Inductive Origin?</th>
<th>Analysis</th>
<th>Interanalyst Agreement for Each Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1995</td>
<td>Bavelas, Chovil, Coates, &amp; Roe (journal article)</td>
<td>No</td>
<td>a. Training data&lt;br&gt;b. One speaker in a dyad&lt;br&gt;c. One speaker who was alone&lt;br&gt;<strong>Experiment 2</strong>&lt;br&gt;d. Locating <em>topical</em> vs. <em>interactive gestures</em>&lt;br&gt;e. <em>Degree of redundancy of gesture with words</em> in its phonemic clause (0 to 3)&lt;br&gt;Do interactive gestures occur more in dialogue than in alternating monologues, and do they elicit predictable responses from addressees?&lt;br&gt;<strong>Study 1:</strong>&lt;br&gt;a. Locating <em>topical</em> vs. <em>interactive gestures</em>&lt;br&gt;<strong>Study 2:</strong> Identifying <em>addressee’s first</em> response after an interactive gesture:&lt;br&gt;b. Six possible responses; e.g., <em>confirming</em> (nod, “mhmm”), <em>specific content, hesitation</em></td>
<td>a. All five analysts ≥90%&lt;br&gt;b. Median pairwise = 90%&lt;br&gt;(range = 75% to 100%)&lt;br&gt;c. Median pairwise = 88%&lt;br&gt;(range = 71% to 94%)&lt;br&gt;d. Median pairwise = 86%&lt;br&gt;(range = 83% to 95%)&lt;br&gt;e. Two analysts, two checks:&lt;br&gt;r = .93, .95</td>
</tr>
<tr>
<td>1995</td>
<td>Johnson (master’s thesis)</td>
<td>Yes</td>
<td>How do dyads make decisions about ambiguous stimuli?&lt;br&gt;a. Dividing utterances into units with <em>only one function</em>&lt;br&gt;b. Identifying the <em>specific function</em> of each unit, i.e., <em>hypothesis</em>, <em>individual fact</em>, or <em>social fact</em></td>
<td>a. Four analysts; pairwise agreements ≥ 90%&lt;br&gt;b. Three analysts: overall pairwise agreement = 83%</td>
</tr>
<tr>
<td>1995</td>
<td>Pinch (doctoral dissertation)</td>
<td>No</td>
<td>Does neurological status (right, left, or no hemispheric damage) affect spontaneous irony in dialogue?&lt;br&gt;a. Identifying each utterance as <em>ironic</em> or <em>not ironic</em>&lt;br&gt;b. Locating <em>ironic utterances only</em></td>
<td>a. 87%&lt;br&gt;b. 90%&lt;br</td>
</tr>
<tr>
<td>Year</td>
<td>Study</td>
<td>Inductive Origin?</td>
<td>Analysis</td>
<td>Interanalyst Agreement for Each Analysis</td>
</tr>
<tr>
<td>------</td>
<td>-------------------------------</td>
<td>-------------------</td>
<td>--------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>1996</td>
<td>Schindler (honor's thesis)</td>
<td>Yes</td>
<td>Do situations that are not happy (e.g., when arguing or guessing wrong) elicit smiles?</td>
<td>a. 85% over all tasks (range = 82% to 100%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>a. Locating <em>smiles</em> during not-happy situations within five different tasks</td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td>Bavelas, Coates, &amp; Johnson (journal article)</td>
<td>Yes</td>
<td>What is the nature of the responses addressees make when they are listening to a close-call story?</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Experiment 1:</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>a. Deciding whether to divide a contiguous listener response or not</td>
<td>a. &gt;90%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>b. Distinguishing between <em>generic</em> listening responses (e.g., “mhm”) and <em>specific</em> listener responses (e.g., wincing)</td>
<td>b. 95%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>c. Global rating of the quality of the <em>story plot</em></td>
<td>c. Intraclass r (7 raters) = .69</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>d. Global rating of the <em>quality of the story telling</em></td>
<td>d. Intraclass r (3 raters) = .83</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>e. Number of negative features of the story ending (e.g., <em>appropriate</em>, <em>abrupt</em>, <em>choppy</em>, or <em>overjustified</em>)</td>
<td>e. Two analysts vs. a third analyst, $r = .76$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Experiment 2:</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>f. Distinguishing between <em>generic</em> and <em>specific listener responses</em></td>
<td>f. 99.6%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>g. Number of negative features of the story ending (e.g., <em>appropriate</em>, <em>abrupt</em>, <em>choppy</em>, or <em>overjustified</em>)</td>
<td>g. Two analysts vs. a third analyst, $r = .69$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Note: The analysis excluded smiles when they occurred alone.</td>
<td></td>
</tr>
<tr>
<td>2002</td>
<td>Bavelas, Coates, &amp; Johnson (journal article)</td>
<td>Yes</td>
<td>What predicts the placement of the listener responses found in Bavelas et al. (2000)?</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>a. Identifying periods of <em>mutual gaze</em></td>
<td>a. Two dyads: 100%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>b. Timing the <em>onset and offset of mutual gaze</em></td>
<td>b. First dyad: 84% within 0.1 second</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Second dyad: 87% within 0.2 seconds</td>
</tr>
</tbody>
</table>

*(Continued)*
<table>
<thead>
<tr>
<th>Year</th>
<th>Study</th>
<th>Inductive Origin?</th>
<th>Analysis</th>
<th>Interanalyst Agreement for Each Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005</td>
<td>Woods (honor’s thesis)</td>
<td>No</td>
<td>Do hand gestures show a given-new effect, becoming abbreviated over repeated trials with the same partner?</td>
<td>a. 100%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>a. Whether the speaker <em>gestured</em> during a given trial</td>
<td>b. 96%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>b. Whether the gesture was <em>analyzable</em></td>
<td>c. $r = .995$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>c. The <em>duration</em> of each gesture</td>
<td></td>
</tr>
<tr>
<td>2004;</td>
<td>Tomori (honor’s thesis); Tomori &amp; Bavelas</td>
<td>No</td>
<td>Do experts in two different approaches to psychotherapy differ in how they talk with clients?</td>
<td>Separately for each of the four therapists:</td>
</tr>
<tr>
<td>2007</td>
<td>(journal article)</td>
<td></td>
<td>a. Whether the therapist’s utterance was a <em>formulation</em>, a <em>question</em>, <em>both</em>, or <em>neither</em></td>
<td>a. 80%, 86%, 95%, and 95%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>b. Whether the content of a formulation or question was <em>positive</em>, <em>negative</em>, <em>both</em>, or <em>neither</em></td>
<td>b. 85%, 86%, 90%, and 100%</td>
</tr>
<tr>
<td>2007</td>
<td>Phillips (doctoral dissertation)</td>
<td>Yes</td>
<td>Experiment 2: How do face-to-face dialogues differ from online chats or bulletin-board formats?</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>a. Identifying a single, discrete <em>idea unit</em> (topic)</td>
<td>a. 90%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>b. Identifying <em>topically related idea units</em></td>
<td>b. 87%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>c. Separating idea units into <em>statements, replies, questions, and answers</em></td>
<td>c. 88%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>These combined to measure:</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>o Development of ideas: the proportion of replies to each other’s statements</td>
<td>d. First two analysts: 96%; third analyst: 95%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>o Coherence: proportion of replies that were not adjacent to the original statement and proportion of answers that were not adjacent to the question</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>o Number of <em>audible listener responses</em> (used to measure the density of listener responses: the ratio of listener responses to words spoken)</td>
<td></td>
</tr>
<tr>
<td>Year</td>
<td>Study</td>
<td>Inductive Origin?</td>
<td>Analysis</td>
<td>Interanalyst Agreement for Each Analysis</td>
</tr>
<tr>
<td>------</td>
<td>----------------------------------------------------------------------</td>
<td>------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>------------------------------------------</td>
</tr>
<tr>
<td>2008</td>
<td>Bavelas, Gerwing, Sutton, &amp; Prevost (journal article)</td>
<td>Yes</td>
<td>Does being in a dialogue or monologue or being visible to the addressee (or not) affect the rate or kind of hand gestures?</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>a. Whether each hand movement was <em>gesturing</em></td>
<td>a. 92%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>b. Whether each period of <em>gesturing should be divided</em> (i.e., locating individual gestures)</td>
<td>b. 78%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>c. Total number of <em>gestures per speaker</em></td>
<td>c. $r = .98$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>d. <em>Bias check:</em> each speaker's rate of gesturing by naive and nonnaive analysts</td>
<td>d. $r = .89$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>e. Whether the function of each gesture was <em>topical, interactive, or picture oriented</em></td>
<td>e. 92%; 93%; $\kappa = .70$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>f. Total number of <em>topic gestures per speaker</em></td>
<td>f. $r = .996$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>g. Total number of <em>interactive gestures per speaker</em></td>
<td>g. $r = .87$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>h. Total number of <em>picture-oriented gestures per speaker</em></td>
<td>h. $r = .93$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>i. Size of each gesture (scale of 1 to 5)</td>
<td>i. $r = .92$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>j. Average size of gesture per speaker</td>
<td>j. $r = .97$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>k. Whether there was a <em>verbal deictic</em> with each topic gesture</td>
<td>k. 97%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>l. Proportion of <em>topic gestures with verbal deictics per speaker</em></td>
<td>l. $r = .98$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>m. Whether each topic gesture for two specific features was <em>redundant with the words that accompanied the gesture vs. conveying additional information</em></td>
<td>m. First feature: 84%; second feature: 83%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>n. Proportion of each speaker's <em>topic gestures that were redundant with words</em></td>
<td>n. First feature: $r = .92$; second feature: $r = .71$ ($r = .95$ without the highest and lowest outliers)</td>
</tr>
<tr>
<td>Year</td>
<td>Study</td>
<td>Inductive Origin</td>
<td>Analysis</td>
<td>Interanalyst Agreement for Each Analysis</td>
</tr>
<tr>
<td>------</td>
<td>--------------------------------------</td>
<td>------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>------------------------------------------</td>
</tr>
<tr>
<td>2008</td>
<td>Gerwing (doctoral dissertation)</td>
<td>Yes</td>
<td>Using home videos, was an infant who was later diagnosed with autism less socially responsive to his parents’ social overtures than his two triplet siblings were? Study 1 (on sample of home videos)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>a. All three infants = 98%; infant A = 96%; B = 98%; C = 98%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>a. <em>Is each infant on screen</em> (second by second)?</td>
<td>b. All three infants = 96%; infant A = 97%; B = 94%; C = 97%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>b. <em>Is a parent directing an overture (i.e., a social action) toward one infant</em> (yes or no, second by second)?</td>
<td>c. All three infants = 81%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>c. Is a parent <em>directing an overture toward one infant</em> (not based on time, which is potentially inflated, but on episodes of social actions)?</td>
<td>d. Source = 100%; function = 98%; modality = 89%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>d. <em>Should the overture be divided</em> (based on source, function, or modality)?</td>
<td>e. All three infants = 85%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>e. Did the infant make an observable response to the overture?</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Study 2 (on the home videos not included in Study 1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(See sample decision tree in Appendix 9.B.)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>a. <em>Is a parent directing an overture</em> (i.e., a social action) toward one infant who is onscreen (yes or no, second by second)?</td>
<td>a. 97%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>b. Same as above, calculated by <em>change from overture to not overture</em> (i.e., event by event)</td>
<td>b. 84%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>c. Who was the agent of the overture (<em>mother, father, both</em>)?</td>
<td>c. 100%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>d. Should the overture be divided, based on source, function, or modality?</td>
<td>d. 85%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>e. What was the function of the overture: <em>attention-seeking, directing, helping, greeting, playful, conversational, rewarding, narrative, or instrumental</em></td>
<td>e. 85%</td>
</tr>
<tr>
<td>Year</td>
<td>Study</td>
<td>Inductive Origin?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>-------</td>
<td>-------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2009</td>
<td>Gerwing &amp; Allison (journal article)</td>
<td>No</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

What are the differences between three methods for analyzing the relationship of hand gestures to words?

- a. Distinguishing between periods of gesturing, holding a gesture, and not gesturing
  - a. 87%
- b. Whether the gesture was about their task (creating a floor plan) or not
  - b. 91%
- c. Whether the gesture was about an identifiable room in the floor plan or not
  - c. 80%
- d. Whether the words in each speech-gesture combination included a deictic expression (e.g., “this” or “here”) or not
  - d. 96%
- e. Whether the gesture depicted a room vs. an object within a room
  - e. 94%
- f. Whether the gesture or words conveyed information about each of four semantic features
  - f. 95%
- g. Whether the information in the words and the gesture was general vs. specific
  - g. 90%
<table>
<thead>
<tr>
<th>Year</th>
<th>Study</th>
<th>Inductive Origin?</th>
<th>Analysis</th>
<th>Interanalyst Agreement for Each Analysis</th>
</tr>
</thead>
</table>
| 2011 | Bavelas, Gerwing, Allison, & Sutton (book chapter) | Yes | Do addressees understand gestures that are not redundant with speech, as shown by completed grounding sequences?  
   a. Locating gestures depicting an identifiable room in the floor plan  
   b. Whether each gesture was redundant with words or not  
   c. Whether the addressee’s response to a nonredundant gesture was explicit vs. implicit  
   d. Whether the addressee’s explicit response was positive vs. negative (e.g., indicated understanding vs. asked for clarification)  
   e. Whether the addressee’s implicit response was positive vs. moot  
   f. Whether the speaker/gesturer’s acknowledgment was explicit, implicit, or other | a. Range = 80% to 97%  
   b. 97%  
   c. 89%  
   d. 96%  
   e. 91%  
   f. 87% |
| 2011 | Gerwing & Allison (journal article) | No | Do speakers shift semantic information between words and hand gestures as a function of visibility?  
   a. Whether the speaker’s words conveyed information in each of five categories  
   b. Whether the speaker’s gestures conveyed information in each of the five categories | a. 92%  
   b. 86% |
| 2013 | Healing (master’s thesis) | Yes | In oncology consultations, do patients contribute information about their individual perspective (i.e., patient-centered information) as well as biomedical information?  
   a. Whether a patient utterance contributed small talk, a generic listening response, biomedical information, or patient-centered information | a. 90% |

(See decision tree in Appendix 9.B.)
<table>
<thead>
<tr>
<th>Year</th>
<th>Study</th>
<th>Inductive Origin?</th>
<th>Analysis</th>
<th>Interanalyst Agreement for Each Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>2013</td>
<td>Korman, Bvelas, &amp; De Jong (journal article)</td>
<td>Yes</td>
<td>What are the ways in which therapists’ formulations change what their clients have said?</td>
<td>a. 90%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>a. Whether each speaking turn by a therapist included a formulation or not</td>
<td>b. 96%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>b. Locating the exact words in the utterance that were the formulation</td>
<td>c. 90%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>c. Identifying each word in the client’s previous utterances that was preserved exactly, preserved deictically, or omitted in the formulation</td>
<td>d. 89%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>d. Identifying words in the formulation that preserved the client’s words in altered form (e.g., a paraphrase); words that were not in what the client said (i.e., added by the therapist); and discourse markers (e.g., “So you mean . . . ”)</td>
<td></td>
</tr>
<tr>
<td>2013</td>
<td>Smock Jordan, Froerer, &amp; Bvelas (journal article)</td>
<td>No</td>
<td>Do therapists differ in their use of positive vs. negative topical content as a function of their model of therapy?</td>
<td>a. For the six therapists, 75% to 88%; median = 85%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>a. Whether each therapist utterance contained positive content, negative content, both, or neither</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Does the positive vs. negative topical content of the therapist affect the next client utterance?</td>
<td>b. For the six clients, 69% to 87%, median = 80%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>b. Whether each client utterance contained positive content, negative content, both, or neither</td>
<td></td>
</tr>
<tr>
<td>2014b</td>
<td>Bvelas, Gerwing, &amp; Healing (journal article)</td>
<td>No</td>
<td>Does being in a dialogue (vs. a monologue) increase the rate of demonstrations, independently of the effect of visibility?</td>
<td>a. 91%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>a. Identifying the speakers’ exact words that accompanied a facial portrayal of a character in the movie they were retelling</td>
<td></td>
</tr>
</tbody>
</table>

(Continued)
<table>
<thead>
<tr>
<th>Year</th>
<th>Study</th>
<th>Inductive Origin?</th>
<th>Analysis</th>
<th>Interanalyst Agreement for Each Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>2014c</td>
<td>Bavelas, Gerwing, &amp; Healing (book chapter)</td>
<td>No</td>
<td>b. Identifying events in the movie in which the speaker quoted a character in the movie</td>
<td>b. 97%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>c. Identifying the exact words in each quotation</td>
<td>c. 99%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>d. Locating speakers' hand gestures</td>
<td>d. 94%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>e. Deciding whether to divide periods of continuous gesturing</td>
<td>e. 93%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>f. Locating speakers' nouns or noun phrases referring to any of nine features in the picture they were describing</td>
<td>f. 94%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>g. Deciding whether each reference was literal vs. figurative</td>
<td>g. 98%</td>
</tr>
<tr>
<td>2014</td>
<td>Gerwing &amp; Dalby (journal article)</td>
<td>Yes</td>
<td>Do participants shift information into visible means (i.e., facial as well as hand gestures) when having a face-to-face dialogue vs. on the phone?</td>
<td>a. 93%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>a. Deciding whether speakers contributed information about each of three semantic features using visible vs. audible means</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Do physicians use semantic gestures when speaking with patients during conversations about treatment plans?</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>a. Decide whether physicians' hand movements are gestures.</td>
<td>a. 100%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>b. Decide whether physicians' gestures have semantic, beat, or interactive functions</td>
<td>b. 91% (Cohen’s $\kappa = .82$)</td>
</tr>
</tbody>
</table>
Appendix 9.B

Examples of Decision Trees

Decision Tree from Healing (2013)

1. Locate patient utterance
   - Is this utterance small talk? Is this the kind of information that you might give someone when you meet them for the first time? E.g., talking about the weather or where you were born?
     - yes → Small talk
     - no → Is this utterance a generic listener response? Is the patient just showing that he or she understands or is following what the physician or family member has said? E.g., saying things like "Yeah" or "Mhm."
       - yes → Generic response
       - no → Is the utterance a question? Is it asking for information or implicitly asking the physician to confirm that the patient is understanding?
         - yes → Patient centered
         - no → Is the information in the utterance biomedically centered? Does it contain only biomedical or medical procedural information (e.g., about medical tests, appointments, or information from other physicians) without any explicit indication from the patient whether the illness, treatment, side effects, or symptoms are (1) affecting the patient’s life, (2) interfering with the patient’s activities, or (3) tolerable to the patient?
           - yes → Biomedical
           - no → Is this utterance patient centered? Does the utterance contain biomedical or medical procedural information with an explicit indication from the patient of whether or how the illness, treatment, side effects, or symptoms are either (1) affecting the patient’s life, (2) interfering with the patient’s activities, or (3) tolerable to the patient? OR does the utterance include information about the patient’s hopes, dreams, plans, goals, preferences, decision-making preferences, wishes, concerns, or fears?
             - yes → Patient centered
             - no → Other
Decision Tree from Gerwing (2008)

Summary of response analysis (please use in conjunction with detailed rules):

1. In the video clip, select the time from a couple of seconds before overtire onset up to the actual time of onset (i.e., not including overtire). Watch a few times.
2. Note baby's behavior during this section on your sheet.
3. Now watch again but allow selection to play past overtire onset into the overtire, but not past onset of next overtire.
4. Now use this decision tree:

   Once the overtire starts, is the baby sufficiently visible for you to judge his or her behavior?
   
   - Yes
   - No

   Write "not visible" in the "is there a change" cell and go on to the next overtire.

   Does the clip end or transition to another section right away before you can judge the baby's behavior?
   
   - No
   - Yes

   If you see an obvious response, note the response. If you don't, write "no time" in the "is there a change" cell and go on to the next overtire.

   If there is another overtire to that baby listed on the sheet, does it start before you can judge the baby's behavior?
   
   - No
   - Yes

   In the immediately previous overtire (if there is one), does the agent make the same proposal?
   
   - Yes
   - No

   If you see an obvious response, note the response. If you don't, write "no time" in the "is there a change" cell and go on to the next overtire.

   Does the baby's behavior change in one of the following ways:
   
   - baby starts a new behavior
   - baby stops current behavior
   - baby increases ongoing activity
   - baby decreases ongoing activity

   - Yes
   - No

   Write "same" in behavior description cell and "N" in the "is there a change" cell and go on to the next overtire.

   Is the timing of the change of behavior at or after the onset of the overtire?
   
   - Yes
   - No

   Write "N" in the "is change predictable" cell and go on to the next overtire.

   Note new behavior and write "Y" in the "is there a change" cell.

   Is there any way you can consider the change predictable, given the nature of the overtire?
   
   - Yes
   - No

   Write "Y" in the "is change predictable" cell and go on to next overtire.
Decision Tree from Bavelas et al. (1992)

Paraphrase the gesture, and examine its form closely.

Does the gesture convey an unambiguous CONTENT meaning?

yes

Is the form in any way INTERACTIVE?

no

TOPICAL

yes

Is it possible to divide this into two gestures?

no

Resolving by balancing the following:
Is there a clear paraphrase for both meanings? Which paraphrase fits the form better (less ambiguous)?

Clues
- Examine the precise words, including stress, interrogative intonation, and immediately preceding context.
- Is the person looking at his or her gesture or at the other person?
- Is it clearly interactive in form (not slightly down, locating something, or moving)?
- At this point, do the best you can!

yes

INTERACTIVE

no

INTERACTIVE

no

TOPICAL