Clark, H. H., and Brennan, S. A. (1991).
In L.B. Resnick, J.M. Levine, & S.D. Teasley (Eds.).
Perspectives on socially shared cognition.
Washington: APA Books.

CHAPTER 7

GROUNDING IN COMMUNICATION

HERBERT H. CLARK AND SUSAN E. BRENNAN

GROUNDING

It takes two people working together to play a duet, shake hands, play chess, waltz, teach, or make love. To succeed, the two of them have to coordinate both the content and process of what they are doing. Alan and Barbara, on the piano, must come to play the same Mozart duet. This is coordination of content. They must also synchronize their entrances and exits, coordinate how loudly to play forte and pianissimo, and otherwise adjust to each other's tempo and dynamics. This is coordination of process. They cannot even begin to coordinate on content without assuming a vast amount of shared information or common ground—that is, mutual knowledge, mutual beliefs, and mutual assumptions (Clark & Carlson, 1982; Clark & Marshall, 1981; Lewis, 1969; Schelling, 1960). And to coordinate on process, they need to update their common ground moment by moment. All collective actions are built on common ground and its accumulation.

We thank many colleagues for discussion of the issues we take up here. The research was supported in part by National Science Foundation Grant BNS 83-20284 and a National Science Foundation Graduate Fellowship.

Correspondence concerning this chapter should be addressed to Herbert H. Clark, Department of Psychology, Jordan Hall, Building 420, Stanford University, Stanford, CA 94305-2130, or Susan E. Brennan, Department of Psychology, State University of New York at Stony Brook, Stony Brook NY 11794-2500.

Communication, of course, is a collective activity of the first order. When Alan speaks to Barbara, he must do more than merely plan and issue utterances, and she must do more than just listen and understand. They have to coordinate on content (Grice, 1975, 1978). When Alan refers to "my dogs," the two of them must reach the mutual belief that he is referring to his feet. They must also coordinate on process. Speech is evanescent, and so Alan must try to speak only when he thinks Barbara is attending to, hearing, and trying to understand what he is saying, and she must guide him by giving him evidence that she is doing just this. Accomplishing this, once again, requires the two of them to keep track of their common ground and its moment-by-moment changes.

In communication, common ground cannot be properly updated without a process we shall call *grounding* (see Clark & Schaefer, 1987, 1989; Clark & Wilkes-Gibbs, 1986; Isaacs & Clark, 1987). In conversation, for example, the participants try to establish that what has been said has been understood. In our terminology, they try to ground what has been said—that is, make it part of their common ground. But how they do this changes a great deal from one situation to the next. Grounding takes one shape in face-to-face conversation but another in personal letters. It takes one shape in casual gossip but another in calls to directory assistance.

Grounding is so basic to communication—indeed, to all collective actions—that it is important to understand how it works. In this chapter we take up two main factors that shape it. One is *purpose*—what the two people are trying to accomplish in their communication. The other is the *medium* of communication—the techniques available in the medium for accomplishing that purpose, and what it costs to use them. We begin by briefly describing grounding as it appears in casual face-to-face conversation. We then consider how it gets shaped by other purposes and in other media.

Grounding in Conversation

What does it take to contribute to conversation? Suppose Alan utters to Barbara, "Do you and your husband have a car?" In the standard view of speech acts (e.g., Bach & Harnish, 1979; Searle, 1969), what Alan has done is *ask* Barbara whether she and her husband have a car, and, in this way, he has carried the conversation forward. But this isn't quite right. Consider this actual exchange:[1]

[1]All examples, except those marked otherwise, come from the so-called London-Lund corpus (Svartvik & Quirk, 1980). We retain the following symbols from the London-Lund notation: "." for a brief pause (of one light syllable); "–" for a unit pause (of one stress unit or foot); "," for the end of a tone unit, which we mark only if it comes mid-turn; "(laughs)" or single parenthesis for contextual comments; "((words))" or double parentheses for incomprehensible words; and "*yes*" or asterisks for paired instances of simultaneous talk.

Alan:	Now, - um, do you and your husband have a j- car
Barbara:	- have a car?
Alan:	Yeah
Barbara:	No -

Even though Alan has uttered "Do you and your husband have a car?", he hasn't managed to ask Barbara whether she and her husband have a car. We know this because Barbara indicates, with "- have a car?", that she hasn't understood him.[2] Only after Alan has answered her query (with "yeah") and she is willing to answer the original question ("no -") do the two of them apparently believe he has succeeded. So asking a question requires more than uttering an interrogative sentence. It must also be established that the respondent has understood what the questioner meant.

Of course, understanding can never be perfect. We assume that the criterion people try to reach in conversation is as follows (Clark & Schaefer, 1989; Clark & Wilkes-Gibbs, 1986): The contributor and his or her partners mutually believe that the partners have understood what the contributor meant to a criterion sufficient for current purposes. This is called the *grounding criterion*. Technically, then, grounding is the collective process by which the participants try to reach this mutual belief. To see some of the forms grounding takes in conversation, let us consider the process of contributing to conversation. Here we will follow a model proposed by Clark and Schaefer (1989) that was founded on a long tradition of work on turns and repairs by Sacks, Schegloff, Jefferson, and others (e.g., Sacks, Schegloff, & Jefferson, 1974; Schegloff, Jefferson, & Sacks, 1977; Schegloff, 1982).

Contributing to Conversation

Most contributions to conversation begin with the potential contributor presenting an utterance to his or her partner. In our example, Alan presents Barbara with the utterance, "Now, - um do you and your husband have a j- car." Why does he present it? Because he wants Barbara to hear it, register it, and understand what he means by it. But he cannot know whether he has succeeded unless she provides evidence of her understanding. In our example, indeed, she provides evidence that she has *not* understood him yet. It is only after the exchange, "- have a car?" and "yeah," that she gives positive evidence of understanding by initiating the answer "no." So contributing to conversation generally divides into two phases:

[2]Actually, the word *ask* is ambiguous between "utter an interrogative sentence" and "succeed in getting the addressee to recognize that you want certain information." Note that you can say, "Ken asked Julia 'Are you coming' but failed to ask her whether she was coming because she couldn't hear him." We will use *ask* in the second sense.

Presentation phase: A presents utterance *u* for B to consider. He does so on the assumption that, if B gives evidence *e* or stronger, he can believe that she understands what he means by *u*.
Acceptance phase: B accepts utterance *u* by giving evidence *e* that she believes she understands what A means by *u*. She does so on the assumption that, once A registers that evidence, he will also believe that she understands.

It takes both phases for a contribution to be complete.

The presentation phase can become very complicated. One way is by self-repairs. In our example, Alan doesn't present the pristine utterance, "Do you and your husband have a car," but rather the messier, "now, - um do you and your husband have a j- car." He expects Barbara to see, for example, that "j-" isn't part of the sentence he is ultimately committed to. Establishing what he is and is not ultimately committed to is no easy task. Another complication is embedding. The presentation itself can contain distinct contributions each with its own presentation and acceptance phases (we will see examples of embedding later in this chapter).

Grounding becomes most evident in the acceptance phase. By the end of A's presentation of some utterance *u*, the partner B may believe she is in one of these states for all or part of *u*:

State 0: B didn't notice that A uttered any *u*.
State 1: B noticed that A uttered some *u* (but wasn't in state 2).
State 2: B correctly heard *u* (but wasn't in state 3).
State 3: B understood what A meant by *u*.

In our example, Barbara apparently thinks she is in state 3 for the first part of Alan's presentation but in state 2 for the final phrase. Because she wants to be in state 3 for the entire presentation, she needs to clear up her understanding of the final phrase. This is what leads her to initiate the *side sequence* (Jefferson, 1972) or *insertion sequence* (Schegloff, 1972) with "- have a car?" All of this is part of the acceptance phase, and so Alan's contribution divides up this way:

Presentation phase:
Alan: Now, - um do you and your husband have a j-car
Acceptance phase:
Barbara: - have a car?
Alan: Yeah

Actually, the acceptance phase only gets completed when Barbara initiates the answer "no" and Alan accepts it as the evidence he needs.

The acceptance phase may also contain embedded contributions. Barbara's "- have a car?" is the presentation phase of a contribution that is wholly contained within the acceptance phase of the main contribution. It is accepted when Alan says "yeah," which is itself a presentation with its own acceptance (see Clark & Schaefer, 1987). So contributions often emerge in hierarchies. They may contain contributions embedded within both their presentation and their acceptance phases.

There is an essential difference, therefore, between merely uttering some words—a presentation—and doing what one intends to do by uttering them—a contribution. When you say to a friend, "I want you to meet Mr. Jones," it isn't guaranteed that you have succeeded in introducing him to Mr. Jones. His hearing aid may have been off. He may have misheard you. Or he may have misunderstood you, as Chico did Groucho in this exchange:

Groucho: Ravelli, I want you to meet Mr. Jones.
Chico: Awright, where should I meet him?

Even without Chico around, grounding is essential.

Evidence In Grounding

Once we utter something in a conversation, one might suppose, all we need to look for is *negative evidence*—evidence that we have been misheard or misunderstood. If we find some, we repair the problem, but if we don't, we assume, by default, that we have been understood. This is, indeed, what is explicitly or tacitly assumed in many accounts of language use (e.g., Grosz & Sidner, 1986; Litman & Allen, 1987; Stalnaker, 1978). When Barbara says "- have a car?" she is giving Alan negative evidence and a clue to what she has misunderstood. But if negative evidence is all we looked for, we would often accept information we had little justification for accepting. In fact, people ordinarily reach for a higher criterion. As the contribution model says, people ultimately seek *positive evidence* of understanding. Let us look at the three most common forms of positive evidence and see how they work.

First, *acknowledgments* are the most obvious form of positive evidence. By acknowledgments we mean much of what has been called *back-channel responses*. These include continuers such as *uh huh*, *yeah*, and the British *m* (Schegloff, 1982), as in the following example:

B: Um well I ha((dn't)) done any English at *all,*
A: *((1 syll))*
B: You know, since O-level.
A: Yea .
B: And I went to some second year seminars, where there are only about half a dozen people,
A: *m*
B: *and* they discussed what ((a)) word was,
A: **m**
B: **and -** what's a sentence, that's *ev*en more difficult.

A: *yeah* yeah -
(and so on)

Continuers are used by partners, according to Schegloff, to signal that they are passing up the opportunity to initiate a repair on the turn so far and, by implication, that they think they have understood the turn so far. Acknowledgments also include assessments, such as *gosh*, *really*, and *good God* (see Goodwin, 1986), and gestures such as head nods that have much the same force as continuers (see Goodwin, 1981). Acknowledgments are generally produced without the speaker taking a turn at talk.

A second, common form of positive evidence is the initiation of the *relevant next turn*. Consider this exchange:

A: Did you know mother had been drinking -
B: I don't think, mother had been drinking at all .

Suppose A is trying to ask B a question. If B understands it, she can be expected to answer it in her next turn. Questions and answers form what are called *adjacency pairs*, and once the first part of an adjacency pair is on the floor, the second part is conditionally relevant as the next turn (Schegloff & Sacks, 1973). So A looks for B to provide not just any utterance, but an answer to his question. If B's utterance is appropriate as an answer, as in our example, it is also evidence that she has understood A's question. If it is not appropriate, it is evidence that she has not understood A's question, as caricatured here:

Miss Dimple: Where can I get a hold of you?
Chico: I don't know, lady. You see, I'm very ticklish.
Miss Dimple: I mean, where do you live?
Chico: I live with my brother.

Chico's answer gives Miss Dimple evidence that he has misunderstood her question, and that leads to the correction in her following turn (for an authentic, spontaneous example, see Clark & Schaefer, 1989). So B may initiate the next turn as positive evidence of her understanding, but A will not take it that way unless it shows her understanding to be correct.

What makes a next turn appropriate or relevant? That isn't difficult to decide for the second part of an adjacency pair—the answer to a question, the response to a request, or the acceptance of an invitation. It also isn't difficult to decide for most other next turns. Conversation generally divides into coherent sections that have identifiable entries, bodies, and exits (see, e.g., Schegloff & Sacks, 1973). These sections are devoted to one or another social process, such as making plans or exchanging information. Most turns are designed to carry that process forward and give evidence about the speaker's understanding of the previous step in the process. As Sacks et al. (1974, p. 728) noted, "Regularly, then, a turn's talk will display its speaker's understanding of a prior turn's talk, or whatever other

talk it marks itself as directed to." Initiating the relevant next turn is ordinarily an excellent piece of positive evidence.

Requiring positive evidence of understanding seems to lead to an infinite regress. The problem is this: When B says "uh huh" or "you're there all day" in response to A's presentation, she herself is making a presentation. Now her presentation, being more words, requires positive evidence of understanding from A, which requires him to give more words. But his words constitute another presentation that she must accept with more words, and so on ad infinitum. If every presentation were accepted with positive evidence in the form of words, the process would spin out to infinity. Empirically, it is easy to show that people do not take an infinite number of words to contribute to a conversation. How, then, *do* they do it?

There is no infinite regress in the contribution model because some forms of evidence, such as the relevant next turn and continued attention (our next form of evidence), do not have separate presentations. A relevant next turn provides positive evidence of understanding of the presentation phase it follows, but it does so by initiating the next contribution without a break. So although the acceptance process can spin out for many turns, it usually ends with the partner initiating a relevant next turn. Take the following example, in which A is presenting a book identification number:

A: F .six two
B: F six two
A: Yes
B: Thanks very much

The first presentation is accepted with a repetition, the repetition with an acknowledgment, and the acknowledgment with a thanks, which is the next contribution at the level of original contribution.

The third and most basic form of positive evidence is *continued attention*. In conversation people monitor what their partners are doing moment by moment—in particular, what they are attending to. If Alan presented an utterance while Barbara wasn't paying attention, he could hardly assume that she was understanding him. She must show that she is paying attention, and one way is through eye gaze. Suppose she is looking away from Alan. As Goodwin (1981) has shown, Alan can try to capture her gaze—and also, presumably, her attention—by starting an utterance. Just as she begins turning to him, he will start the utterance over again. Or he can start an utterance, pause until she starts turning, and then go on with the utterance. Speakers have many ways of getting a partner's attention.

Positive evidence of understanding comes with attention that is unbroken or undisturbed. Alan has reason to believe Barbara is following him as long as she continues to attend in the expected way. Whenever she turns to listen to someone else, looks puzzled, or hangs up the telephone, Alan has reason to believe

that he has lost her. She is no longer hearing or understanding to criterion. Ordinarily, that will push him into taking corrective action.

Least Collaborative Effort

People apparently don't like to work any harder than they have to, and in language use this truism has been embodied in several principles of least effort. The traditional version exhorts the speaker: Don't expend any more effort than you need to get your addressees to understand you with as little effort. Grice (1975) expressed this idea in terms of two maxims: *Quantity*—Make your contribution as informative as is required for the current purpose of the exchange, but do not make your contribution more informative than is required—and *Manner*—Be brief (avoid unnecessary prolixity). According to both versions, speakers are supposed to create what we will call *proper utterances*, ones they believe will be readily and fully understood by their addressees.

The principle of least effort, however, assumes flawless presentations and trouble-free acceptances. It does not allow for grounding and, therefore, cannot do justice to what really happens in conversations. Here are just three problems with this principle (Clark & Wilkes-Gibbs, 1986).

1. *Time pressure*. Speakers appear to limit the time and effort they allow for planning and issuing each utterance, and that often leads them to issue improper utterances. They may utter a sentence or phrase, discover it to be inadequate, and then amend it, as in "Number 7's the goofy guy that's falling over—with his leg kicked up." They may start a phrase, think better of it, and start a different phrase, as in "We must ha- we're big enough to stand on our own feet now." They may create patently improper parts of utterances, such as *what's his name* in "If he puts it into the diplomatic bag, as um - what's his name, Micky Cohn did, . then it's not so bad." They may invite their interlocutors to complete their utterances, as in this exchange (Wilkes-Gibbs, 1986):

A: That tree has, uh, uh
B: Tentworms.
A: Yeah.
B: Yeah.

If all these speakers had taken the time and effort needed, they could have produced proper utterances—flawless performances. They didn't. The principle of least effort says that they should have.

2. *Errors*. Speakers often issue improper utterances because they make errors and have to repair them, as in Alan's "now, - um do you and your husband have a j- car?" If Alan had taken more time and effort, he could have avoided errors and dysfluencies. Why didn't he?

3. *Ignorance*. Speakers sometimes realize they just don't know enough about their interlocutor to design a proper utterance, so they are forced to issue an improper utterance instead. Take the person who was trying to identify an abstract figure that resembled an ice skater (Clark & Wilkes-Gibbs, 1986), who said, "Um, the next one's the person ice skating that has two arms?" Why the question intonation, or what Sacks and Schegloff (1979) have called a *try marker*? With it the speaker was indicating that he was not sure that his definite description "the person ice skating that has two arms" was adequate to pick out the right figure. He was asking his partner whether it was adequate and, if it was not, inviting an alternative description. Here, no matter how hard the speaker tried, he might not have managed a proper utterance. So why did he do what he did?

The principle of least effort, it has been argued, must therefore be replaced with the following principle (Clark & Wilkes-Gibbs, 1986):

> *The principle of least collaborative effort*: In conversation, the participants try to minimize their collaborative effort—the work that both do from the initiation of each contribution to its mutual acceptance.

Such a principle helps account for many phenomena. Consider repairs. As Schegloff et al. (1977) noted, speakers have two strong preferences about repairs: (a) They prefer to repair their own utterances rather than let their interlocutors do it, and (b) they prefer to initiate their own repairs rather than let their interlocutors prompt them to do it. Although these two preferences have many causes, the upshot is that they minimize collaborative effort. As for preference 1, it generally takes less effort for the speaker than for an interlocutor to make a repair. An interlocutor will usually need extra turns, and he or she has to get the speaker to accept the repair anyway. As for preference 2, it usually takes less effort for the speaker than for an interlocutor to initiate a repair. The interlocutor will generally create extra turns in doing so, whereas the speaker will not. Every extra turn adds to collaborative effort.

Also, speakers often realize that it will take more collaborative effort to design a proper utterance than to design an improper utterance and enlist their addressees' help. Speakers, for example, can present a provisional utterance and add try markers to ask for confirmation. They can present a difficult utterance in installments and check for understanding after each installment (as we will describe later). They can invite addressees to complete an utterance they are having trouble with. And they have many other collaborative techniques at their disposal. The principle of least collaborative effort is essential for a full account of face-to-face conversation.

GROUNDING CHANGES WITH PURPOSE

People in conversation generally try to establish collective purposes (Grice, 1975). If they are planning a party, that may be their overall collective purpose. In each

section of the conversation, their purpose might be to complete pieces of that plan, and in each subsection it would be even more specific. Other times their overall purpose might be to get acquainted, swap gossip, or instruct and learn. Grounding should change with these purposes. If addressees are to understand what the speaker meant "to a criterion sufficient for current purposes," then the criterion should change as their collective purposes change. So, too, should the techniques they exploit. Techniques should change, for example, with the content of the conversation—with what needs to be understood. Indeed, specialized techniques have evolved for grounding different types of content. We will illustrate with two types of content—references and verbatim content.

Grounding References

Many conversations focus on objects and their identities; when they do, it becomes crucial to identify the objects quickly and securely. Conversations like these arise, for example, when an expert is teaching a novice how to build things, and the two of them refer again and again to pieces of the construction. They arise in court when lawyers and witnesses try to establish the identities of persons, places, and things. They also arise in tasks in which people have to arrange figures, post cards, blocks, color chips, or other such objects. In psychology, an entire industry has been built on this type of task patterned after Krauss and Weinheimer's (1964, 1966, 1967) original referential communication task. Yet conversations like these are common enough in real life.

The purpose of interest here is to establish *referential identity*—that is, the mutual belief that the addressees have correctly identified a referent. There are several common techniques for establishing this.

1. *Alternative descriptions.* When speakers refer to objects, they typically use one or more referring expressions—a definite or indefinite description, proper noun, demonstrative, or pronoun. One way their partners can demonstrate that they have identified the referent or can check on its identity is by presenting an alternative description, as in this interchange:

> A: Well, that young gentleman from - ((the park)) .
> B: Joe Joe Wright you mean? - - *(- - laughs)*
> A: *yes, (laughs) yes*
> B: ((God)), I thought it was old Joe Wright who(('d)) walked in at first

A describes a referent as "that young gentleman from the park"; B gives evidence of having identified the man by offering an alternative description; he adds the question intonation to get confirmation of that description; and A accepts that description. This technique is common whenever referential identity is at stake (e.g., Clark & Wilkes-Gibbs, 1986; Isaacs & Clark, 1987).

2. *Indicative gestures.* When a speaker refers to a nearby object, the partners can give positive evidence that they have identified it by pointing, looking, or touching. In this example, S had been handed a photograph of a flower patch (Clark, Schreuder, & Buttrick, 1983):

> B: How would you describe the color of this flower?
> S: You mean this one [pointing]?
> B: Yes.
> S: It's off yellow.

S confirmed the referent of B's "this flower" by pointing.

3. *Referential installments.* It is often important to establish the identity of a referent before saying something about it. The reason is simple. Until the referent has been properly identified, the rest of the utterance will be difficult, if not impossible, to understand. The speaker can secure the reference by treating it as an installment of the utterance to be confirmed separately. Take this exchange between an expert and a novice assembling a pump (Cohen, 1984):

> S: Take the spout—the little one that looks like the end of an oil can—
> J: Okay.
> S: and put that on the opening in the other large tube. With the round top.

In the first line, S presents "the spout—the little one that looks like the end of an oil can" and then pauses for evidence that J has identified the referent. He goes on only when that installment has been grounded.

In English, there is a specialized construction for just this purpose called *left-dislocation*. It is traditionally illustrated with invented examples such as *Your dog he just bit me*. This example begins with a "left dislocated" noun phrase, *your dog*, followed by a full sentence with a pronoun, *he*, referring to the same object. In genuine conversation, left-dislocation rarely looks like this. A more typical example is this second exchange from S and J:

> S: Okay now, the small blue cap we talked about before?
> J: Yeah.
> S: Put that over the hole on the side of that tube—
> J: Yeah.
> S: —that is nearest to the top, or nearest to the red handle.

As Geluykens (1988) has shown, 29% of the left dislocated noun phrases in the London-Lund corpus (Svartvik & Quirk, 1980) are followed by an intervening move from the interlocutor (either a continuer or something more extensive), as in our example. Another 52% are followed by a pause during which

the partner could have nodded acceptance. Left-dislocation may have evolved for just this specialized purpose—grounding references separately.

4. *Trial references.* Speakers can also initiate the grounding process for a reference in mid-utterance. When speakers find themselves about to present a name or description that they are not sure is entirely correct or comprehensible, they can present it with a *try marker* followed by a slight pause, and get their partners to confirm or correct it before completing the presentation. Consider this example:

A: So I wrote off to . Bill, . uh who ((had)) presumably disappeared by this time, certainly, a man called Annegra?
B: Yeah, Allegra.
A: Allegra, uh replied, . uh and I . put . two other people [continues].

A apparently wants to assert, "A man called Annegra replied, and I . . .". But being uncertain about the name *Annegra*, he presents it with a try marker. B confirms, with "yeah," that she knows who he is referring to, but then corrects the name to *Allegra*. A accepts the correction by re-presenting *Allegra* and continuing. The entire correction is made swiftly and efficiently.

There are other techniques adapted for this purpose, but the four we have mentioned give an idea of their range and specialization.

Grounding Verbatim Content

Sometimes it is important to register the verbatim content of what is said. When a friend tells you a telephone number, you do more than listen for the gist of it. You try to get it verbatim so that you can copy it down or rehearse it until you dial the number. The same goes for names, addresses, book titles, credit card numbers, bank accounts, dollar amounts, and library call numbers. These are specialized situations, and specialized grounding techniques have evolved for them. Here are a few:

1. *Verbatim displays.* When customers call directory enquiries for a telephone number, they often confirm the number they are given with a verbatim display, as in this British example (Clark & Schaefer, 1987):

O: It's Cambridge 12345
C: 12345
O: That's right.
C: Thank you very much.

C confirms the number that O has presented him by repeating it verbatim, "12345." In the British calls studied by Clark and Schaefer (1987), customers responded to the operators' number presentations with verbatim displays over

70% of the time. Operators, in turn, often responded to the customers' presentations of names, towns, and street addresses with verbatim displays.

2. *Installments.* When speakers present a lot of information to be registered verbatim, they generally cut it up into bite-sized chunks, or installments, and receive verbatim displays on each installment, as in this example:

A: Ah, what ((are you)) now, *where*
C: *yes* forty-nine Skipton Place
A: Forty-one
C: Nine . nine
A: Forty-nine, Skipton Place,
C: W one .
A: Skipton Place, . W one, ((so)) Mr D Challam
C: Yes
A: Forty-nine Skipton Place, W one,
C: Yes
A: Right oh.

C divides his address into repeatable chunks, and A gives a verbatim display for each. Speakers seem able to divide most types of information into such chunks. They do it spontaneously, for example, for recipes presented over the telephone (Goldberg, 1975).

Dividing a presentation into repeatable installments is based on the tacit recognition that people have limited immediate memory spans. Even the telephone company recognizes this and divides telephone numbers into conventional installments of three or four digits in size. In the calls to directory enquiries studied by Clark and Schaefer (1987), British operators always divided numbers of seven or more digits into their conventional groupings.

3. *Spelling.* For many words, getting the verbatim content right means getting the spelling right. So contributors often spell out critical words, as in the following:

A: The name is Iain, . I A I N .
C: m
A: Lathom-Meadows, that's L A T, . H O M, . hyphen, - Meadows,
C: Yes .

Or they have other tricks for getting the spelling right, as illustrated here:

B: And my name, is James Persian-Omo, that's Persian like the carpet
C: Yes
B: Hyphen, . Omo like the detergent, O M O
C: Yeah.

Other times, it is the partners who do the spelling as they confirm a name.

On occasion, two partners in conversation will set different criteria, one stricter than the other. Imagine a father instructing a 5-year-old son on how to play a game or work a machine. The son may think he understands an instruction while the father still has serious doubts. The father may go on testing for understanding long after the son thinks he needs to.

To summarize, specialized techniques have evolved for grounding special pieces of conversation. When it is critical that a reference be well established, people will use techniques that are custom designed for that purpose. When it is the verbatim content that is crucial, they will use other techniques. In this way, grounding changes with the current purpose.

GROUNDING CHANGES WITH THE MEDIUM

By the principle of least collaborative effort, people should try to ground with as little combined effort as needed. But what takes effort changes dramatically with the communication medium. The techniques available in one medium may not be available in another, and even when a technique is available, it may cost more in one medium than in the other. Our prediction is straightforward: People should ground with those techniques available in a medium that lead to the least collaborative effort.

Consider the acknowledgment *okay*. In face-to-face or telephone conversations, it can be timed precisely so that it constitutes evidence of understanding and not an interruption. In keyboard teleconferencing—when people communicate over keyboards and screens—it is difficult to time an acknowledgment precisely, and trying to do so may interrupt the other typist. So the cost of an acknowledgment is higher in this medium.

Media come in a great variety, and new ones are being introduced every year. Think of the telegraph, videotape, picturephone, express mail, fax machines, electronic bulletin boards, and little yellow post-it notes. Here we will consider a sample of two-way personal media: face-to-face conversation, the telephone, video teleconferencing, keyboard teleconferencing, answering machines, electronic mail (email), and personal letters. For now we will put aside such one-way, broadcastable media as books, newspapers, television, and radio. Some personal media have been compared experimentally (see, for example, Ochsman & Chapanis, 1974), but most of these studies identify and describe differences without a theoretical framework for explaining them. We propose to set them in a framework that will account for many of their differences.

Constraints on Grounding

Personal media vary on many dimensions that affect grounding. Here are eight constraints that a medium may impose on communication between two people, A and B.

1. *Copresence: A and B share the same physical environment.* In face-to-face conversation, the participants are usually in the same surroundings and can readily see and hear what each other is doing and looking at. In other media there is no such possibility.

2. *Visibility: A and B are visible to each other.* In face-to-face conversation, the participants can see each other, and in other media they cannot. They may also be able to see each other, as in video teleconferencing, without being able to see what each other is doing or looking at.

3. *Audibility: A and B communicate by speaking.* Face to face, on the telephone, and with some kinds of teleconferencing, participants can hear each other and take note of timing and intonation. In other media they cannot. An answering machine preserves intonation, but only some aspects of utterance timing.

4. *Cotemporality: B receives at roughly the same time as A produces.* In most conversations, an utterance is produced just about when it is received and understood, without delay. In media such as letters and electronic mail, this is not the case.

5. *Simultaneity: A and B can send and receive at once and simultaneously.* Sometimes messages can be conveyed and received by both parties at once, as when a hearer smiles during a speaker's utterance. Simultaneous utterances are also allowed, for example, in the keyboard teleconferencing program called *talk*, where what both parties type appears letter by letter in two distinct halves of the screen. Other media are cotemporal but not simultaneous, such as the kind of keyboard teleconferencing that transmits characters only after the typist hits a carriage return.

6. *Sequentiality: A's and B's turns cannot get out of sequence.* In face-to-face conversation, turns ordinarily form a sequence that does not include intervening turns from different conversations with other people. With email, answering machines, and letters, a message and its reply may be separated by any number of irrelevant messages or activities; interruptions do not have the same force.

7. *Reviewability: B can review A's messages.* Speech fades quickly, but in media such as email, letters, and recorded messages, an utterance stays behind as an artifact that can be reviewed later by either of the partners—or even by a third party. In keyboard teleconferencing, the last few utterances stay visible on the screen for awhile.

8. *Revisability: A can revise messages for B.* Some media, such as letters and email, allow a participant to revise an utterance privately before sending it to a partner. In face-to-face and telephone conversations, most self-repairs must be done publicly. Some kinds of keyboard teleconferencing fall in between; what a person types appears on the partner's screen only after every carriage return, rather than letter by letter.

Table 1
SEVEN MEDIA AND THEIR ASSOCIATED CONSTRAINTS

Medium	Constraints
Face-to-face	Copresence, visibility, audibility, cotemporality, simultaneity, sequentiality
Telephone	Audibility, cotemporality, simultaneity, sequentiality
Video teleconference	Visibility, audibility, cotemporality, simultaneity, sequentiality
Terminal teleconference	Cotemporality, sequentiality, reviewability
Answering machines	Audibility, reviewability
Electronic mail	Reviewability, revisability
Letters	Reviewability, revisability

There are other differences across media, but these are among the most important for grounding. Table 1 characterizes seven personal media by these constraints.

Costs of Grounding

When a medium lacks one of these characteristics, it generally forces people to use alternative grounding techniques. It does so because the costs of the various techniques of grounding change. We will describe eleven costs that change. The first two, formulation and production costs, are paid by the speaker. The next two, reception and understanding costs, are paid by the addressee. The rest are paid by both. We emphasize that these costs are not independent of each other.

Formulation costs

It costs time and effort to formulate and reformulate utterances. It costs more to plan complicated than simple utterances, more to retrieve uncommon than common words, and more to create descriptions for unfamiliar than familiar objects. It costs more to formulate perfect than imperfect utterances. As we will see, these costs are often traded off for others, depending on the medium.

Production costs

The act of producing an utterance itself has a cost that varies from medium to medium. It takes little effort (for most of us) to speak or gesture, more effort to type on a computer keyboard or typewriter, and the most effort (for many of us, anyway) to write by hand. Speaking is swift, typing is slower, and handwriting is slowest. These costs are traded off for other costs as well. People are willing to use more words talking than in typewriting to accomplish a goal, and the faster people are at typing, the more words they are willing to use.

Reception costs

Listening is generally easy, and reading harder, although it may be easier to read than to listen to complicated instructions or abstract arguments. It also costs to have to wait while the speaker produces a turn. This wait takes its toll in keyboard conversations when addressees must suffer as they watch an utterance appear letter by letter with painstaking backspacing to repair misspellings.

Understanding costs

It is also more costly for people to understand certain words, constructions, and concepts than others, regardless of the medium. The costs can be compounded when contextual clues are missing. Email, for example, is neither cotemporal nor sequential. That makes understanding harder because the addressee has to imagine appropriate contexts for both the sender and the message, and to remember what the message is in response to, even when the "subject" field of the message is filled in.

Start-up costs

This is the cost of starting up a new discourse. It is the cost of getting B initially to notice that A has uttered something and to accept that he or she has been addressed. Start-up costs are minimal face to face, where A need only get B's attention and speak. They are a bit higher when A must get to a telephone, look up a number, dial it, and determine that the answerer is B. They are often higher yet in email. First, A has to get access to the right software and hardware, find the right email address, and start the message. Second, the message may not reach the addressee if the channel is unreliable or the address has typos in it. Third, depending on the system, the sender may or may not be notified of its delivery. And finally, once the message is delivered, there is no guarantee that the addressee will read it right away. There are similar start-up costs in writing letters.

Delay costs

These are the costs of delaying an utterance in order to plan, revise, and execute it more carefully. In face-to-face conversation, as in all cotemporal and simultaneous media, these costs are high because of the way delays, even brief delays, are interpreted. When speakers leave too long a gap before starting a turn, they may be misheard as dropping out of the conversation or as implying other more damaging things. And when they leave too long a pause in the middle of a turn, they may be misheard as having finished their turn. With the pressure to minimize

both midturn pauses and preturn gaps, speakers are often forced to utter words they may have to revise or to let their addressees help them out. Even when it is clear that a delay is due to the speaker's production difficulty, it costs the addressees to wait. In media without cotemporality—such as email and personal letters—delays that would be crippling in conversation are not even noticeable, and so their costs are nil. But in cotemporal media, the cost of a delay can be high: When the drugged Juliet failed to respond, Romeo did himself in.

Delay costs often trade off with formulation costs. In writing letters, we can take our time planning and revising each sentence. Computerized text editing has made this even easier. But in face-to-face and telephone conversations, where delay costs are high, we have to formulate utterances quickly. That forces us to use simpler constructions and to be satisfied with less than perfect utterances. The other media lie between these two extremes.

Asynchrony costs

In conversation, people time their utterances with great precision (Jefferson, 1973). They can begin an utterance precisely at the completion of the prior speaker's turn. They can time acknowledgments to mark what it is they are acknowledging. They can interrupt a particular word to show agreement or disagreement on some aspect of it. In media without copresence, visibility, audibility, or simultaneity, timing is much less precise, and without cotemporality, it is altogether impossible. So grounding techniques that rely on precision of timing should go up in cost when production and reception are asynchronous.

Speaker change costs

In conversation the general rule is, "Two people can speak at the same time only for short periods or about limited content." The rule is usually simplified to "One speaker at a time" (see Sacks et al., 1974), and it tends to hold for other media as well. But the cost of changing speakers varies with the medium. In face-to-face conversation it is low. The participants find it easy to arrange for one speaker to stop and another to start. There are regular rules for turn taking in which the points of possible change in speakers are frequent, easily marked, and readily recognized, and the changes can be instantaneous. Also, the costs of simultaneous speech, at least for short intervals and limited content, are minor. The participants usually continue to understand without disruption.

The cost of changing speakers is higher in media with fewer cues for changes in turns. Costs are quite high, for example, in keyboard teleconferencing, where the points of speaker change are not as easily marked or readily recognized. These points may need to be marked by a convention such as the use of "o" (for "over"), a device that is also used by airplane pilots and citizens' band radio operators, when only one party can be heard at a time. Speaker change costs are greater still in letters, answering machines, and email, where it may take much

work for one participant to stop and another to start up. Changing the speaker in these media is a little like starting up a communication from scratch. One effect of high speaker change costs is that people try to do more within a turn.

Display costs

In face-to-face conversation, it is easy to point to, nod at, or present an object for our interlocutors. It is also easy to gaze at our interlocutors to show them we are attending, to monitor their facial expressions, or to pick them out as addressees. In media without copresence, gestures cost a lot, are severely limited, or are out of the question. In video teleconferencing, we can use only a limited range of gestures, and we cannot always look at someone as a way of designating "you." Showing pictures is possible with media such as video, fax machines, and letters.

Fault costs

There are costs associated with producing an utterance fault, that is, any mistake or missaying. Some faults lead to failures in understanding, and failures in one utterance are likely to undermine the next one. The costs of these faults increase with the gravity of the failures. Other faults make the speaker look foolish, illiterate, or impolite, so they also have their costs. The cost of most faults trades off with what it costs to repair them (our next category) or to prevent them in the first place. To avoid paying fault costs, speakers may elect to pay more in formulation costs. But it depends on the medium. In conversation, a hearer may expect faults from a speaker because the production of speech is so spontaneous. In email, faults are not as easily justified, because the sender has already had a chance to revise them, and because the damage done is not as easily repaired.

Repair costs

Some repairs take little time or effort; others take a lot, and still others are impossible to make. Because faults tend to snowball, speakers should want to repair them as quickly as possible. In audible conversation, as we noted earlier, speakers prefer to initiate and make their own repairs, and there is evidence that they interrupt themselves and make these repairs just as soon as they detect a fault (Levelt, 1983). These preferences tend to minimize the cost of a repair. Self-corrections take fewer words and turns than do repairs by others, and so do repairs initiated by oneself rather than by others. These preferences also help minimize the cost of faults: They tend to remove a fault from the floor as quickly as possible. In media that are not cotemporal, repairs initiated or made by others become very costly indeed, so speakers will try hard to avoid relying on others to repair misunderstandings. It is less costly for them to revise what they say before sending it. Another way to minimize repair costs may be to change to a different medium to make the repair.

Cost Tradeoffs

People manage to communicate effectively by all the media we have mentioned, but that does not mean that they do so in the same way in each medium. The way people proceed reflects the costs they incur. Recall the intermediate states 0 to 3 mentioned earlier for face-to-face conversation: (0) A and B failing to establish any connection yet, (1) A getting B's attention, (2) B perceiving A's utterance correctly, and (3) B understanding what A meant. There are costs to getting to each state, and in some media the states are quite distinct. In media that are not cotemporal, there is the additional problem that A does not have immediate evidence as to which of these states B is in with respect to A's utterance. In a medium such as email, B's lack of response can be highly ambiguous. Did she not get the message, did she get it and not read it, did she read it and choose not to respond, did she not understand it, or what? A does not know whether B is in state 0, 1, 2, or 3.

Once we assume that people need to ground what they say and that they trade off on the costs of grounding, we can account for some of the differences in language use across media. Consider a study reported by Cohen (1984) in which tutors instructed students on assembling a pump. Their communication was either by telephone or by keyboard. Over the telephone, tutors would first get students to identify an object, and only when they had confirmed its identification did they ask students to do something with it.

> S: Uh, now there's a little plastic blue cap.
> J: Yep.
> S: Put that on the top hole in the cylinder you just worked with.

In contrast, in keyboard conversations, tutors would identify an object and instruct students what to do with it all in a single turn.

> K: Next, take the blue cap that has the pink thing on it and screw it to the blue piece you just screwed on.

That is, there were many more separate requests for identification over the telephone than in keyboard conversations. According to Cohen, "Speakers attempt to achieve more detailed goals in giving instructions than do users of keyboards" (Cohen, 1984, p. 97).

But why? The principle of least effort suggests a reason: The two media have different profiles of grounding costs. Over the telephone, it doesn't cost much to produce an utterance or change speakers. On a keyboard, it costs much more. So to minimize these costs, tutors and students on a keyboard might seek and provide evidence after larger constituents; that is, they should try to do more within each turn than they would over the telephone. This prediction would account for the difference Cohen describes between the two media. In addition, repairs are more costly over the keyboard; if tutors and students are aware of this, they might formulate their utterances more carefully. But because delay costs can be just as high over the keyboard as on the telephone, and speaker change costs are higher, there are likely to be more misunderstandings over a keyboard, and repairs will take more collaborative effort. The differences among these and other media can be explained by the techniques participants choose for grounding. They balance the perceived costs for formulation, production, reception, understanding, start-up, delay, timing, speaker change, display, faults, and repair.

Medium and Purpose Interact

Grounding techniques depend on both purpose and medium, and these sometimes interact. Face-to-face conversations appear to be preferred for reprimanding, whereas telephone conversations or letters may be preferred for refusing an unreasonable request (Furnham, 1982). In a study of working groups, face-to-face conversation was preferred for negotiating and reaching consensus, whereas email was preferred for coordinating schedules, assigning tasks, and making progress reports (Finholt, Sproull, & Kiesler, 1990).

These preferences can be explained in terms of the costs associated with each medium relative to the participants' purposes. Sometimes the participants want a reviewable record of a conversation—as for schedules, task assignments, and progress reports; and other times they do not. Sometimes speakers want to get a hearer's full attention, and sometimes they want to avoid interrupting. Sometimes people want their reactions to be seen, as in negotiating and reaching consensus, and sometimes they do not. Which medium is best for which purpose, then, depends on the form grounding takes in a medium and whether that serves the participants' purposes.

Finally, as more participants join the conversation and the medium must support the work of a whole group, costs and tradeoffs shift. Start-up costs may be greater. Reception costs will increase if a hearer must put effort into identifying who is speaking or writing. Fault costs and repair costs will be higher when a group is involved. Any medium that supports cooperative work can be evaluated in terms of the techniques it allows for grounding.

CONCLUSION

Grounding is essential to communication. Once we have formulated a message, we must do more than just send it off. We need to assure ourselves that it has been understood as we intended it to be. Otherwise, we have little assurance that the discourse we are taking part in will proceed in an orderly way. For whatever

we say, our goal is to reach the grounding criterion: that we and our addressees mutually believe that they have understood what we meant well enough for current purposes. This is the process we have called *grounding*.

The techniques we use for grounding change both with purpose and with medium. Special techniques have evolved, for example, for grounding references to objects and for grounding the verbatim content of what is said. Grounding techniques also change with the medium. In the framework we have offered, media differ in the costs they impose on such actions as delaying speech, starting up a turn, changing speakers, making errors, and repairing errors. In grounding what we say, we try to minimize effort for us and our partners. Ordinarily, this means paying as few of these costs as possible.

The lesson is that communication is a collective activity. It requires the coordinated action of all the participants. Grounding is crucial for keeping that coordination on track.

References

Bach, K., & Harnish, R. M. (1979). *Linguistic communication and speech acts*. Cambridge, MA: MIT Press.

Clark, H. H., & Carlson, T. (1982). Hearers and speech acts. *Language, 58*(2), 332–373.

Clark, H. H., & Marshall, C. R. (1981). Definite reference and mutual knowledge. In A. K. Joshi, B. L. Webber, & I. A. Sag (Eds.), *Elements of discourse understanding* (pp. 10–63). Cambridge, England: Cambridge University Press.

Clark, H. H., & Schaefer, E. F. (1987). Collaborating on contributions to conversations. *Language and Cognitive Processes, 2*(1), 19–41.

Clark, H. H., & Schaefer, E. F. (1989). Contributing to discourse. *Cognitive Science, 13*, 259–294.

Clark, H. H., Schreuder, R., & Buttrick, S. (1983). Common ground and the understanding of demonstrative reference. *Journal of Verbal Learning and Verbal Behavior, 22*, 1–39.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition, 22*, 1–39.

Cohen, P. R. (1984). The pragmatics of referring and the modality of communication. *Computational Linguistics, 10*(2), 97–146.

Finholt, T., Sproull, L., & Kiesler, S. (1990). Communication and performance in ad hoc task groups. In J. Galegher, R. Kraut, & C. Egido (Eds.), *Intellectual teamwork: Social and technological foundations of cooperative work* (pp. 291–326). Hillsdale, NJ: Erlbaum.

Furnham, A. (1982). The message, the context, and the medium. *Language and Communication, 2*, 33–47.

Geluykens, R. (1988). The interactional nature of referent-introduction. *Papers from the 24th Regional Meeting, Chicago Linguistic Society*, 141–154.

Goldberg, C. (1975). A system for the transfer of instructions in natural settings. *Semiotica, 14*, 269–296.

Goodwin, C. (1981). *Conversational organization: Interaction between speakers and hearers*. New York: Academic Press.

Goodwin, C. (1986). Between and within: Alternative sequential treatments of continuers and assessments. *Human Studies, 9*, 205–217.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics, Volume 3: Speech acts* (pp. 225–242). New York: Seminar Press.

Grice, H. P. (1978). Some further notes on logic and conversation. In P. Cole (Ed.), *Syntax and semantics, volume 9: Pragmatics* (pp. 113–128). New York: Academic Press.

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

Isaacs, E. A., & Clark, H. H. (1987). References in conversation between experts and novices. *Journal of Experimental Psychology, 116*, 26–37.

Jefferson, G. (1972). Side sequences. In D. Sudnow (Ed.), *Studies in social interaction* (pp. 294–338). New York: Free Press.

Jefferson, G. (1973). A case of precision timing in ordinary conversation: Overlapped tag-positioned address terms in closing sequences. *Semiotica, 9*, 47–96.

Krauss, R. M., & Weinheimer, S. (1964). Changes in reference phases as a function of frequency of usage in social interaction: A preliminary study. *Psychonomic Study, 1*, 113–114.

Krauss, R. M., & Weinheimer, S. (1966). Concurrent feedback, confirmation, and the encoding of referents in verbal communication. *Journal of Personality and Social Psychology, 4*, 343–346.

Krauss, R. M., & Weinheimer, S. (1967). Effect of referent similarity and communication mode on verbal encoding. *Journal of Verbal Learning and Verbal Behavior, 6*, 359–363.

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition, 14*, 41–104.

Lewis, D. K. (1969). *Convention: A philosophical study*. Cambridge, MA: Harvard University Press.

Litman, D. J., & Allen, J. F. (1987). A plan recognition model for subdialogues in conversation. *Cognitive Science, 11*, 163–200.

Ochsman, R. B., & Chapanis, A. (1974). The effects of 10 communication modes on the behavior of teams during cooperative problem-solving. *International Journal of Man-Machine Studies, 6*, 579–619.

Sacks, H., & Schegloff, E. A. (1979). Two preferences in the organization of reference to persons in conversation and their interaction. In G. Psathas (Ed.), *Everyday language: Studies in ethnomethodology* (pp. 15–21). New York: Irvington.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking in conversation. *Language, 50*, 696–735.

Schegloff, E. A. (1972). Notes on a conversational practice: Formulating place. In D. Sudnow (Ed.), *Studies in social interaction* (pp. 75–119). New York: Free Press.

Schegloff, E. A. (1982). Discourse as an interactional achievement: Some uses of "uh huh" and other things that come between sentences. In D. Tannen (Ed.), *Analyzing discourse: Text and talk. Georgetown University Roundtable on Languages and Linguistics 1981* (pp. 71–93). Washington, DC: Georgetown University Press.

Schegloff, E. A., Jefferson, G., & Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language, 53*, 361–382.

Schegloff, E. A., & Sacks, H. (1973). Opening up closings. *Semiotica, 8*, 289–327.

Schelling, T. C. (1960). *The strategy of conflict*. Oxford: Oxford University Press.

Searle, J. R. (1969). *Speech acts*. Cambridge, England: Cambridge University Press.

Stalnaker, R. C. (1978). Assertion. In P. Cole (Ed.), *Syntax and semantics, Volume 9: Pragmatics* (pp. 315–332). New York: Academic Press.

Svartvik, J., & Quirk, R. (Eds.). (1980). *A corpus of English conversation*. Lund, Sweden: Gleerup.

Wilkes-Gibbs, D. (1986). *Collaborative processes of language use in conversation*. Unpublished doctoral dissertation, Stanford University, Stanford, CA.