

## Common Ground and the Understanding of Demonstrative Reference

HERBERT H. CLARK

*Stanford University*

ROBERT SCHREUDER

*University of Leiden*

SAMUEL BUTTRICK

*Stanford University*

Suppose a speaker gestures toward four flowers and asks a listener, "How would you describe the color of this flower?" How does the listener infer which of the four flowers is being referred to? It is proposed that he selects the one he judges to be most salient with respect to the speaker's and his common ground—their mutual knowledge, beliefs, and suppositions. In a field experiment, it was found that listeners would accept demonstrative references (like *this flower*) with more than one potential referent. Three further experiments showed that listeners select referents based on estimates of their mutual beliefs about perceptual salience, the speaker's goals, and the speaker's presuppositions and assertions. Common ground, it is argued, is necessary in general for understanding demonstrative reference.

A demonstrative reference is a reference that requires an accompanying gesture for its complete interpretation. Suppose Margaret points at a copy of the *New York Times* that Duncan is holding and asks Duncan

(1) Could I look at that newspaper?

To understand what she is referring to, Duncan must not only grasp the words *that newspaper*, but also register what Margaret is indicating by her gesture, or *demonstration*, which could have been a nod, a gaze, a presentation, or some other gesture.

Demonstrative references at first seem trivial to understand. In (1), let us call *newspaper*, the descriptive part of the noun phrase, the *descriptor*, and the newspaper Margaret is

The research reported here was supported in part by Grant MH-20021 from the National Institute of Mental Health and by the Nederlandse Organisatie voor Zuiver-Wetenschappelijk Onderzoek. For helpful suggestions on the research and manuscript, we thank Eve V. Clark, Richard J. Gerrig, Ray W. Gibbs, Jr., Barbara C. Malt, Gregory L. Murphy, and Deanna L. Wilkes-Gibbs. Requests for reprints should be sent to Herbert H. Clark, Department of Psychology, Jordan Hall, Stanford University, Stanford, Calif. 94305.

pointing at the *demonstratum*. Standard theories of demonstrative reference (e.g., Bennett, 1978; Fillmore, 1982; Lakoff, 1974; Lyons, 1975, 1977; Maclaran, 1980) make two tacit assumptions about such a reference. Assumption 1: The referent is identical to the demonstratum. If Margaret is referring to a newspaper, she must be pointing at that newspaper. Assumption 2: The referent is uniquely determined by the demonstratum together with the descriptor. With her gesture, Margaret specifies a set of potential referents—the newspaper itself, the newsprint, a headline, and many other possibilities—and with the descriptor *newspaper*, she uniquely specifies which element in that set, the newspaper itself, is the intended referent. With Assumptions 1 and 2, it seems obvious how people understand demonstrative references.

The problem is that both assumptions are incorrect. Suppose Margaret points at the same newspaper and asserts

(2) I used to work for those people.

Although she is demonstrating a newspaper, she is referring to its publishers, the New York

Times Company. This utterance is perfectly acceptable, and in the right circumstances, Duncan will readily understand it. Contrary to Assumption 1, the referent need not be identical to the demonstratum (Nunberg, 1979). Next, suppose Margaret, gesturing in the direction of two newspapers Duncan is holding, asks Duncan

(3) Could I look at that newspaper?

By Assumption 2, this reference should be unacceptable since there are two newspapers. But if one has a screamer, say "War Over," and the other has ordinary headlines, this utterance is also perfectly acceptable, and Duncan will understand her as referring to the first newspaper. The point is that understanding demonstrative references often requires complicated inferences. In (2), Duncan had to infer the connection between the newspaper being demonstrated and the people being referred to. In (3), he had to infer which of the two newspapers was the referent. Both examples raise the same issue: how do listeners infer the mapping from the demonstratum to the referent?

Our goal is to characterize how people understand demonstrative reference in general. In this paper, however, we will investigate only cases that violate Assumption 2, as in Example (3). There Margaret's reference can be said to be "underdetermined," since her demonstration and the descriptor *newspaper*, by themselves, do not allow Duncan to pick out the intended referent. How, then, does he do it? We will describe a partial model of demonstrative reference (see Clark & Schreuder, Note 1), draw out its consequences for underdetermined reference, and report four experiments illustrating these consequences.

#### DEMONSTRATIVE REFERENCE

In our proposal, a demonstrative reference has three parts: a demonstratum  $d$ , a demonstrative relation  $F$ , and a referent  $r$ . In Example (2),  $d$  is the newspaper Margaret is pointing at;  $r$  is the people who publish it; and  $F$  is the relation of being the people who publish, which we will write *people-who-publish*.

The relation  $F$  maps the demonstratum  $d$  into the referent  $r$  (see Nunberg, 1979). In Example (3), similarly,  $d$  is the set of two newspapers Margaret is pointing at;  $r$  is the one with the screamer; and  $F$  is the relation *having-a-screamer-in*. Example (1) is simply the degenerate case in which the demonstratum  $d$ , the newspaper, coincides with the referent  $r$ , and so  $F$  is the identity relation. In general,  $F$  can be any relation the speaker can get the addressees to infer quickly and uniquely.

Our proposal is that the speaker provides the addressees with enough information about  $d$ ,  $F$ , and  $r$  as a triad that they can infer  $d$ ,  $F$ , and  $r$  uniquely. In (2), Margaret's descriptor *people* narrows  $r$  to people; her demonstration indicates that those people bear some relation to the newspaper; so  $F$  takes the form *people-somehow-connected-with*. Duncan is intended to infer the specific  $d$ ,  $F$ , and  $r$  as the combination, or package, that make the best sense under these constraints. For that, he must use other information and infer  $F$  to be *people-who-publish* and not, say, *people-who-print*, *people-who-distribute*, or *people-mentioned-in-the-headlined-story-in*, which could indeed be the intended  $F$  in other situations.

The method listeners use in packaging  $d$ ,  $F$ , and  $r$ , according to our proposal, changes with the information provided. With a precise demonstration and a vague descriptor, they should choose one method; with a vague demonstration and a precise descriptor, they should use another (see Pechmann & Deutsch, 1982). Ultimately, they should appeal to a general principle they believe to be part of the speaker's and addressees' common ground, the *principle of optimal design*:

The speaker designs his utterance in such a way that he has good reason to believe that the addressees can readily and uniquely compute what he meant on the basis of the utterance along with the rest of their common ground.

Working backwards, the addressee can assume he has been given enough information and can thereby reason through to the speaker's meaning.

The principle of optimal design relies crucially on the notion of *common ground*, tech-

nically the mutual knowledge, beliefs, and assumptions shared by the speaker and addressees (see Clark & Carlson, 1981; Clark & Marshall, 1981). In our proposal, the speaker intends each addressee to base his inferences not on just *any* knowledge or beliefs he may have, but only on their *mutual* knowledge or beliefs—their common ground. One goal of our research is to explore the role common ground plays in demonstrative reference.

#### *Underdetermined Demonstrative Reference*

Let us turn now to underdetermined demonstrative reference. Suppose Julia nods at a cluster of men jogging along a road and tells Ken

(4) That man is my neighbor.

If Ken takes  $d$  to be the set of ten men, and  $r$  to be one of them, what  $F$  should he infer? If one jogger was naked, he might infer  $F$  to be naked-man-in. If exactly one jogger was especially tall, a midget, or in a gorilla suit, or if one had just fallen, won the race, or slugged a bystander, he would infer a different  $F$  and  $r$ . In each case, he would choose  $F$  to make  $r$  the most distinctive man in the cluster. The  $F$  would have the general form most-salient-part-of.

This characterization, however, cannot be complete. Each jogger is the most salient by some criterion—by being the only one with red hair, in second place, or wearing blue socks. In what respect is the *intended* referent most salient? Suppose Julia had just told Ken that her neighbor was completely bald, and suppose all the joggers except one were hirsute. In saying (4), Julia would expect Ken to see that this prior information, part of their common ground, was relevant. He was to pick the jogger most salient not on *general* grounds, but against their *particular* common ground. He was to select the bald man even if the tallest man, the midget, or the winner would be most salient on general grounds. As the principle of optimal design dictates, the only information he should consult is their common ground. So the general form of  $F$  should be most-salient-part-of-with-respect-to-common-ground.

The common ground between two people is based on roughly three sources of information (Clark & Marshall, 1981). The first is *perceptual evidence*, what the two have jointly experienced or are jointly experiencing at the moment. The second is *linguistic evidence*, what the two have jointly heard said or are now jointly hearing as participants in the same conversation. The third is *community membership*. They take as common ground everything they believe is universally, or almost universally, known, believed, or supposed in the many communities and subcommunities to which they mutually believe they both belong. Most parts of common ground are based on a combination of these sources.

Two people's common ground may be a tiny plot or a large acreage. If Julia and Ken are strangers, their common ground comes entirely from their joint membership in the community of, say, adult Americans (mutually recognized from their American accents) and from the scene they are witnessing. If Julia assumed nakedness would be the most salient attribute in the scene for most Americans, she would expect Ken to understand her in (4) as referring to the naked jogger. But if Julia and Ken are intimates, their common ground will be extensive and include their discussion of her bald neighbor, and Ken should understand her as referring to the jogger who is bald. For the same demonstratum and descriptor, the relation  $F$ , and hence the referent  $r$ , should change with common ground.

At first, one might view the underdetermined demonstrative references in (2) through (4) simply as defective, as no different from the equivalent underdetermined definite descriptions. If they were, they should be replaceable by underdetermined definite descriptions with no change in understandability. They are not. Suppose Margaret uttered (1') through (4') in the same contexts as (1) through (4), but without the accompanying gestures.

(1') Could I see the newspaper in your hand?

(2') \*I used to work for the people in your hand.

(3') \*Could I see the newspaper in your hand?

(4') \*The man in the group in front of us is my neighbor.

While most people we have asked find (1') acceptable, they find (2') through (4') unpalatable. They feel as if they are left with the questions "What people?" "Which newspaper?" and "Which man?" whereas they accept (2) through (4) without such questions. When a reference is accompanied by a demonstration, listeners seem to feel they are to understand it partly by examining the demonstratum more closely. Demonstrative references are simply different from other definite descriptions (Hawkins, 1978).

Before examining this *common ground model* further, we need to demonstrate that Assumption 2 is false—that people *will* accept demonstrative references with more than one potential referent. That was the aim of Experiment 1.

#### EXPERIMENT 1

We stopped students on the Stanford University campus, showed them a photograph with four types of flowers in it, and asked, gesturing to the photograph as a whole, "How would you describe the color of this flower?" We expected them to respond in one of three main ways:

- (a) Implicit acknowledgement of understanding, "It's yellow."
- (b) Request for confirmation, "Do you mean this flower [pointing at one of the flowers]?"
- (c) Request for clarification, "Which flower do you mean?"

Response (a) indicates the students thought they understood the reference; (b) indicates less certainty; (c) indicates the least certainty, since the students do not even offer the conjecture they do in (b). Put negatively, (b) indicates they thought the reference was partly defective, and (c), rather more defective. With (a), (b), and (c), we have a natural scale of understandability.

#### Method

The color photograph we used, from the *Sunset New Western Garden Book* (1979, p.

128), depicted four varieties of flowers beside a wooden fence: a cluster of yellow daffodils in the front center; white daisies dispersed throughout the background; a number of orange California poppies at the right; and a cluster of blue irises on the far left. We compared two versions of the photograph. Picture 1 was a color Xerox copy of the photograph, measuring 18 by 24 cm, in which the daffodils were only slightly more prominent than the other three flowers. Picture 2 was a second color Xerox copy, cropped to 11 by 15 cm, in which the daffodils were clearly more salient than the others. The two pictures each showed all four types of flowers but differed in the prominence of the daffodils.

One of us, Buttrick, approached 40 students at various places on the Stanford University campus. With a clipboard in hand, he introduced himself to each student, handed the student one of the two pictures, and, nodding at the picture while preparing to write down the answer, asked, "How would you describe the color of this flower?" He wrote down everything the student said, including all requests for confirmation or clarification. The flower chosen was always clear from the color named or flower pointed at. Each picture was described by 20 students.

#### Results and Discussion

Suppose that listeners look where the speaker is pointing, note the set of potential referents there (the referent array), and select the object uniquely specified by the descriptor. This model, which is much like Olson's (1970) model of definite reference, is based on Assumptions 1 and 2. We will call it the *classical* model. According to that model, if the referent array contains four flowers, listeners should accept a descriptor like *daffodil*, which picks out a unique flower, but they should not accept *flower*, which does not. Yet as shown in Table 1, many students in Experiment 1 accepted *flower* without hesitation. For Picture 2, 11 of 20 students immediately described the color of the daffodils. For the two pictures together, 14 of 40 students did so. To anticipate a finding in Experiment 4, 12 of 15 students gave response (a) for a similar pic-

TABLE 1  
 RESPONSES TO "HOW WOULD YOU DESCRIBE THE COLOR OF THIS FLOWER?" FOR PICTURES 1 AND 2 (EXPERIMENT 1)

| Response            | Picture 1<br>(Low salience) | Picture 2<br>(High salience) |
|---------------------|-----------------------------|------------------------------|
| a. Immediate choice | 3                           | 11                           |
| b. "This one?"      | 3                           | 3                            |
| c. "Which one?"     | 12                          | 5                            |
| Other <sup>a</sup>  | 2                           | 1                            |
| Totals              | 20                          | 20                           |

<sup>a</sup>Two students offered a color for Picture 1 and then questioned their choice; one student narrowed the four referents down to two possibilities.

ture. These responses constitute evidence against the classical model and Assumption 2.

By the classical model, *this flower* should be no more acceptable a reference for Picture 2 than Picture 1, since *flower* picks out a unique flower no more for one than for the other. Yet it was. More students chose the daffodils immediately (response (a)) on Picture 2 than Picture 1, 55 to 15%,  $\chi^2(1) = 7.03$ ,  $p < .01$ . More students offered at least some hypothesis (response (a), (b), or a mixed response) for Picture 2 than Picture 1, 75 to 40%,  $\chi^2(1) = 5.01$ ,  $p < .05$ . And if responses (a), (b), and (c) are assigned confidence values 3, 2, and 1, respectively, students were more confident for Picture 2 than Picture 1 by 2.15 to 1.55,  $t(38) = 2.92$ ,  $p < .01$ . All this evidence also counts against the classical model.

By the common ground model, the more salient the most salient flower is, the more confidently it should be picked as the referent (and the less likely the reference should be deemed defective). On intuitive grounds, the daffodils were more salient and should have been chosen more confidently in Picture 2 than Picture 1, and they were. We will consider perceptual salience of this sort more closely in Experiment 2.

It is surprising how widely the students ranged in their acceptance of the demonstrative references. On each of the two pictures, some students immediately committed themselves to the daffodils (response (a)), yet others wouldn't even offer a conjecture (response (c)). What made the confident students so confident and the uncertain students so un-

certain? If we could answer this, we would be a long way toward the main issue of interest: how do people infer the intended referent when it is underdetermined by the demonstration and descriptor?

#### SALIENCE AND COMMON GROUND

According to the common ground model, the students in Experiment 1 responded (a), (b), or (c) based on their judgment of how salient the most salient flower was. They judged salience, in turn, against their estimate of their common ground with the questioner. Unfortunately, there are currently no theories about how people judge salience against common ground, nor will we propose one here. Instead we propose an empirical measure derived from the *Schelling task*, which we have named after one of the first to use the task, Thomas C. Schelling (1960).

Suppose a student named Mary is shown a picture of three balls—a basketball, a golf ball, and a squash ball—and is told, "Select one of these three balls. I am giving the same picture and instructions to another student in the next room, a person you don't know. You will both get a prize if the two of you select the same ball, but nothing if you don't." As Schelling pointed out (p. 54):

What is necessary is to coordinate predictions, to read the same message in the common situation, to identify the one course of action that their expectations of each other can converge on. They must "mutually recognize" some unique signal that coordinates their expectations of each other.

For this problem, Mary might assume large size to be the most distinctive attribute in her

common ground with the anonymous student. If she takes this as a “unique signal” for coordinating their expectations, she will pick the basketball. Let us call such a choice on a Schelling task a *Schelling choice*.

But suppose Mary is told her partner is Peter, with whom she regularly plays squash. Because she could assume the squash ball was particularly salient in their common ground, she could select it, believing Peter too would see it as a “unique signal” for coordinating expectations. No matter how avid a squash player she was, she could not make the same assumption when her partner was just “another student.” It matters to Mary who her partner is because it matters what is in their common ground.

Schelling choices, therefore, should reflect what the two partners take to be salient with respect to their common ground. In the common ground model, listeners choose referents by the same criterion—salience with respect to common ground—and so the distribution of Schelling choices should predict their choices of referents. It should also predict their confidence. In a Schelling task, if object A is selected more often than object B, then in the corresponding reference task those listeners who happen to select A should be more confident than those who happen to select B. Experiments 2 and 3 were designed to test these predictions.

#### EXPERIMENT 2

Students were asked to interpret demonstrative references from a speaker they did not know. In assessing common ground, therefore, they could make only the most general assumptions about the display and how it would be viewed by most people. In fact, people generally have a pretty good idea of what others will attend to in such a display, especially what they will see as salient, distinctive, or out of place. These estimates should be the same ones they use in making Schelling choices and, more simply, in choosing the most salient object.

#### *Method*

We constructed 27 displays of common objects by cutting pictures out of merchandise

catalogs from such department stores as Sears, J. C. Penney, and Best Products. Each display depicted two to seven objects of the same type, like four watches, six lamps, or three tents, each with a number or letter next to it for identification. The displays were designed so that one object in each display seemed more prominent than the rest—larger, more distinctive in shape, more foregrounded—and so that the prominence of the most prominent object varied from display to display. The 27 displays were placed in one order and photocopied to make three booklets that differed only in the sentence typed at the bottom of each page: “What do you think of this X” for a reference task, “You are both to choose the same X” for a Schelling task, and “You are to choose the most prominent or salient X” for a salience task. For X we substituted a one- or two-word description of the objects in the picture—for example, *watch*, *lamp*, or *tent*.

*Reference task.* Ten students were instructed, “Imagine that you and another person are looking through this catalog. He shows you this page and asks you the question printed at the bottom (for example, ‘What do you think of this alarm clock?’). You are to indicate for each display which object you think he is referring to. After you have made your choice, please indicate how confident you are that you have chosen the correct referent.” They rated their confidence on a 7-point scale, with 1 meaning “no confidence” and 7 “very confident.”

After the students had finished, they were asked to go through the displays again and rank the remaining objects for how likely each one was to be the intended referent. They were then asked to describe for each display the criteria on which they had based their initial choices. Since the reference choices and confidence ratings were collected first, they could not be influenced by the later judgments.

*Schelling task.* Ten other students were instructed, “Imagine that you and another person whom you do not know are looking at these displays. If you can both independently select the same object, you will win a prize. If you fail to choose the same object then you lose.” Otherwise, they followed the same

procedure as in the reference task. They went through the 27 displays one time making choices and confidence ratings, a second time ranking the remaining objects, and a third time describing the criteria for their choices.

*Salience task.* Ten more students were asked to choose for each display "the most prominent or salient *X*." They too went through the displays three times, once making choices and salience ratings, a second time giving ranks, and a third time describing their criteria.

The 30 students were Stanford University undergraduates participating either as a course requirement or for pay. They worked in groups of two to four in sessions lasting about 45 minutes. The first three displays were considered practice and later discarded.

### Results

The choices in the reference task generally coincided with those in the Schelling and salience tasks. In each display for each task, one object tended to be chosen more than any other. Let us call this the major choice. In the reference task, the percentage of students making the major choice averaged 70% (ranging from 40 to 100%). The percentages in the Schelling and salience tasks averaged 81 and 89%, respectively. So within tasks, the students showed considerable agreement. The major choices also tended to coincide across tasks. For 18 of the 24 displays, they were identical for the three tasks; for all 24 displays, they were identical for at least two of the three tasks. For each display we can identify the one object on which the largest number of students agreed across all three tasks. The percentage of students making that choice averaged 67%, which is near the maximum possible of 70%, the average percentage of students making the major choice on the ref-

erence task. So the major choices in the three tasks are in good agreement.

A second way to show agreement across the three tasks is to compare the rankings of the objects in each display. For each task, we computed the mean ranks for the two to seven objects in each display; we then correlated the mean ranks, for each display separately, between the reference and Schelling tasks, between the reference and salience tasks, and between the Schelling and salience tasks. These 72 correlations (3 comparisons for 24 displays) had a mean of .81 and a median of .85. Table 2 shows the means, medians, and ranges of these correlations for the three comparisons separately. The mean correlations for the three comparisons were very similar, ranging from .80 to .84, as were the medians, which ranged from .83 to .89. So this measure, too, is in line with the common ground model: the choices in the reference, Schelling, and salience tasks were very similar.

The students in Experiment 2, as in Experiment 1, ranged widely on most displays in how confident they were in their initial choices, from 1 ("no confidence") to 7 ("very confident"). These ratings yielded a surprising finding: the more popular a given choice turned out to be—without anyone knowing this, of course—the more confident were the people who had made it.

There are several ways of demonstrating this. Consider the reference task. For each display, we can compare the mean confidence on the major choice with the mean confidence on the other choices. For the 17 displays where we could compute both means, the majority choice averaged 4.6 (on the 7-point scale) and the other choices averaged only 3.6,  $F(1,19) = 5.19, p < .05$  (see Clark, 1973). The corresponding two means for the Schell-

TABLE 2  
MEANS, MEDIANS, AND RANGES OF THE 24 CORRELATIONS COMPUTED ON THE MEAN RANKS OF THE STUDENTS' CHOICES BETWEEN THE REFERENCE, SCHELLING, AND SALIENCE TASKS (EXPERIMENT 2)

| Pairs of tasks                | Mean | Median | Range     |
|-------------------------------|------|--------|-----------|
| Reference and Schelling tasks | .80  | .89    | 0.33-1.00 |
| Reference and salience tasks  | .80  | .83    | 0.27-1.00 |
| Schelling and salience tasks  | .84  | .89    | 0.21-1.00 |

ing task were 5.0 and 3.9,  $\min F'(1,23) = 15.20$ ,  $p < .01$ , and for the salience task, 6.1 and 5.2,  $\min F'(1,15) = 7.37$ ,  $p < .025$ .

The same point can be illustrated in another way. For each task, we classified each student's choice on each array by how many of the ten students concurred on that choice, and then computed the mean confidence rating of each of these ten categories (1 to 10 students concurring on the choice). In the reference task, for example, there were four separate displays in which eight students agreed on a single object within the display. These 32 confidence ratings averaged 5.2. The ten mean ratings computed this way for each task are plotted in Figure 1.

In all three tasks, the more students who concurred on a choice, the more confident were the students who made that choice. In the reference, Schelling, and salience tasks, the mean correlations were .79, .91, and .92, respectively. These three coefficients are each significantly greater than zero,  $F(1,9) > 14.94$ ,

$p < .01$ , and they are not significantly different from one another. Figure 1 also shows, for the major Schelling choices only, the mean confidence ratings from the reference task as predicted from the number of students concurring on their Schelling choices. The correlation here, despite a restricted range, was .75.

Although the three tasks should be very similar to each other, the salience task should stand out in one respect. In the reference and Schelling tasks, the students rated their confidence in whether their choice would be the same as the choice of *another* person, someone they did not know. But in the salience task, they rated the salience as they alone judged it, without any implied comparison with another person's judgments. So the students in the reference and Schelling tasks should be less confident that they knew what other people would do than the students in the salience task should be in what they themselves thought. Indeed, for the reference and Schelling tasks,

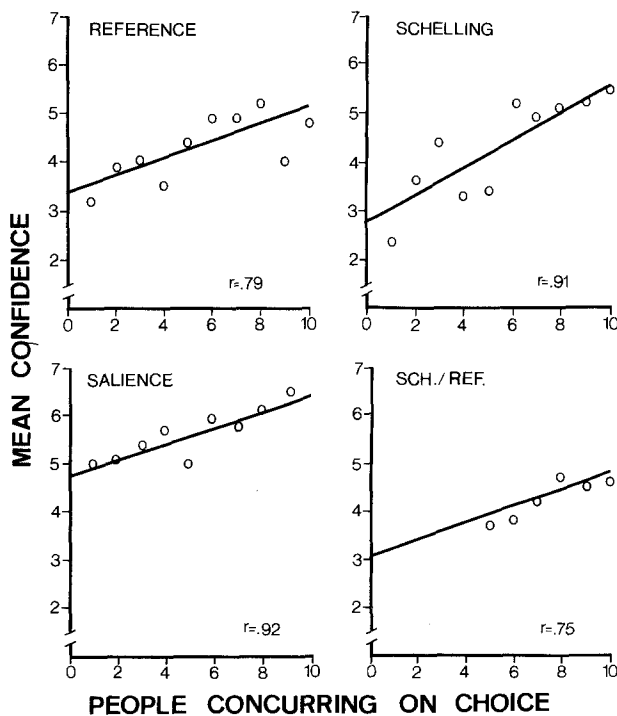


FIG. 1. The mean confidence ratings for choices plotted against number of students concurring on each choice, for the reference, Schelling, and salience tasks separately, and for the reference choices plotted against number of students concurring on the major choices in the Schelling task (Experiment 2).



the mean confidence ratings were 4.4 and 4.8, but for the salience task it was 5.8, which is significantly larger than the other two,  $F(1,27) > 5.76, p < .05$ .

The criteria the students gave for their choices were virtually identical across the three tasks. As representative of all three, here is a breakdown of the 240 reasons for the reference task: 52% mentioned visual features such as size, shape, color, or special markings; 17% mentioned functional properties such as having a carrying case or being simplest to operate; 12% mentioned position in the display, such as being in front; 12% were vague, as in "most unusual," "unique," "different from others," or the like; 4% expressed personal preferences such as "the one I'd be most likely to buy"; and 4% were claims of "don't know" or "no special reason." So about 80% of the reasons mentioned perceptually salient features explicitly; many of the rest could be claims about salience in common ground with people in general. Only one, "My roommate has that kind of basketball," was impossible as a statement about common ground with the speaker.

### Discussion

Although the classical model does not say anything about the relations among reference, Schelling, and salience choices, the common ground model does. Students in the reference task should select in each display that object they think is most salient in their common ground with the speaker. So they should reason in the critical respects just as they would in making a Schelling choice, and they did. The reference choices were well accounted for by the Schelling choices. But the only common ground the students could appeal to in either task was their estimate of how people in general would see the displays—what they would judge to be perceptually salient or prominent. So the reference and Schelling choices should also agree with the salience choices, and they did.

People's confidence in their choice of referents ranged widely on the very same display. What is remarkable is how accurate they

were in estimating their confidence. When John and Mary—pseudonyms for two students—were faced with three clocks in display 9, John chose clock A with a confidence of 6, and Mary chose clock B with a confidence of 3. They presumably made the best choices they could, but just came to different conclusions. Still, each one judged the validity of his or her own choice accurately. John's confidence in clock A was justified, since 80% of the students concurred on this choice. Mary's lack of confidence in clock B was also justified, since no other person agreed with her. So people are quite deft in estimating not merely which object is most likely the intended referent, but also *how* likely it is to be the intended referent. They need both estimates if they are to know when they should ask for confirmation or clarification.

All the findings so far could be accounted for by a *perceptual model* in which the addressee would compare the perceptual salience of all objects in a display and then choose the most salient one. He would not consider the speaker's point of view at all, let alone the common ground they shared. Experiment 3 was designed to test this model.

### EXPERIMENT 3

For most demonstrative references, the common ground contains more than the scene pointed at and assumptions about how it is generally viewed. It also contains what has just been discussed and implied in conversation and, among acquaintances, other information too. In the perceptual model, addressees should ignore all other information and select the perceptually most salient object regardless. In the common ground model, they should select the object most salient against *all* the information in common ground.

In Experiment 3, it was made common ground that the speaker was buying a present either for his old conservative Uncle George or for his young modern Cousin Amanda. In the common ground model, that should make a difference. To a question like "What do you think of this watch?" addressees should select one watch with Uncle George in mind and perhaps quite a different watch for Cousin

Amanda. Perceptual salience should be relevant, but only as weighed against the rest of common ground. By contrast, in the perceptual model the speaker's purpose should make no difference. Addressees should select the same watch regardless of who the speaker was shopping for.

### *Method*

We constructed 21 displays like those in Experiment 2, but each with four objects. Each display was designed so that two objects (say, watches) seemed more appropriate for Cousin Amanda and the other two more appropriate for Uncle George. Again the first three displays were considered practice.

*Reference task.* We prepared several booklets by photocopying the 21 displays with "What do you think of this X?" at the bottom of each display, where X was the appropriate one- or two-word description. Ten students were instructed, "Imagine that your neighbor is looking for a present for his Uncle George. He shows you this catalog and asks 'What do you think of this X?' You are to indicate for each display what object you think your neighbor is referring to." There were ten other students for Cousin Amanda. Uncle George was described as a middle-aged, conservative, thrifty bachelor and Cousin Amanda as a rich, young, modern jet-setter.

As in Experiment 2, the students went through the booklets three times, once making reference choices and confidence ratings, a second time ordering the other objects in each display, and a third time describing criteria for their choices.

*Schelling task.* In this set of booklets, the sentence read "You are both to choose the same X." Ten students were instructed, "Imagine that your Uncle George has looked through this catalog and has chosen one object from each display. Your task is to pick the same object as your Uncle George. If and only if you have both chosen the same object will you win a prize. If you fail to agree in your choice, then you lose." There were 10 other students for Cousin Amanda. Uncle George and Cousin Amanda were described as in the reference task. All 20 students followed the

same procedure as in Experiment 2, giving choices, confidence ratings, rank orders, and rating criteria.

The 40 students, from the same source as in Experiment 2, worked in groups of two to four in sessions lasting about 45 minutes.

### *Results and Discussion*

The reference and Schelling choices for Uncle George were like each other, but unlike those for Cousin Amanda. To show this, we computed the mean ranks of the four objects in each display in each task, and then correlated for each display separately the mean ranks from the four tasks. The means of these correlations are listed in Table 3. If the students made the same choices in the reference and Schelling tasks for the same person, the correlations between these two tasks should be highly positive, and indeed they were. These 18 correlations averaged .75 for Uncle George, and .72 for Cousin Amanda,  $F(1,17) = 87$  and 53, respectively,  $p < .001$ . If the students made roughly complementary choices for Uncle George and Cousin Amanda, the four cross-correlations among the Uncle George and Cousin Amanda tasks should be negative, and they were. The means ranged from  $-.25$  to  $-.44$ , all  $F(1,17) > 5.17$ ,  $p < .05$ , except for  $-.25$ ,  $F(1,17) = 2.45$ .

If the perceptual model were correct, the students should have selected the same perceptually salient object from each display regardless of who the present was for, but they did not. By the common ground model, they should have considered the common ground information about who the present was for, and they did. Their choices in the reference task were predicted by the Schelling choices, but only when the recipient of the gift was the same in the two tasks. So for both choices, the students considered perceptual salience only part of the pertinent information in common ground.

The reasons students gave for their choices were nearly identical for the reference and Schelling tasks, but differed markedly between Uncle George and Cousin Amanda. In the reference task, about 75% of the 360 reasons mentioned visual or functional features,

TABLE 3  
 MEANS OF THE 18 CORRELATIONS COMPUTED ON THE MEAN RANKS OF THE STUDENTS' CHOICES BETWEEN THE REFERENCE  
 AND SCHELLING TASKS WITH UNCLE GEORGE AND COUSIN AMANDA AS RECIPIENTS OF GIFTS  
 (EXPERIMENT 3)

| Task                         | 2     | 3    | 4    |
|------------------------------|-------|------|------|
| 1. Reference task for Amanda | + .72 | -.35 | -.25 |
| 2. Schelling task for Amanda |       | -.44 | -.37 |
| 3. Reference task for George |       |      | +.75 |
| 4. Schelling task for George |       |      | —    |

as in Experiment 2. But unlike the previous experiment, most of these features fit the stereotype for Uncle George or Cousin Amanda, as with "modern design" and "it looks sleek" for Cousin Amanda, and "simplest to use" or "reliable" for Uncle George. Indeed, it is difficult to distinguish these reasons from another 19% that simply identified one object as tailored to one of the two stereotypes, as with "conservative," "plainest," or "fashionable." Another 2% mentioned personal preference, and 4% were vague or "don't know." Only one reason could possibly be typed as egocentric—"It's the only kind I've ever seen." So over 90% of the reasons mentioned features presumed to be in common ground, and most were tailored to the stereotype of who the gift was for.

#### ASSERTIONS AND PRESUPPOSITIONS

One of the surest ways to introduce information into common ground is to mention it in conversation. In Experiment 3, the speaker's goal could have been mentioned in a prior utterance, as in "I am looking for a present for Uncle George—What do you think of this stopwatch?" or in the question itself, as in "What do you think of this stopwatch as a present for Uncle George?" and it probably would have made little difference. But there are many ways of mentioning a piece of information—by asserting or presupposing it, in a prior utterance, the same utterance, or a future utterance, and explicitly or by implication.

A listener must often take account of what is being asserted about the referent. Suppose two men are walking down the street, one very fat and the other very thin, and you say (5),

(6), or (7) to a companion as you nod in their direction.

(5) That man weighs too much for his own good.

(6) That man weighs too little for his own good.

(7) That man is my neighbor.

In (5) and (6), the only way your companion can find a demonstrative relation *F* is by using what you are asserting about the referent; you would assert (5) of the fat man and (6) of the thin man, but not vice versa (unless you were being ironic). In (7), where he can find no reasonable relation, he is uncertain.

The interpretation of the assertion must itself be common ground. Suppose you point at the same two men but say instead

(8) That is what George will look like very soon.

If you had been talking about George gaining weight, your companion would take you as referring to the fat man as in (5). If you had been talking about George losing weight, he would select the thin man as in (6). Without such information, he would be uncertain. In (8), then, your companion appeals to the common ground specifically established between him and you. An overhearer not privy to your earlier conversation would be uncertain regardless of what your companion thought (see Clark & Carlson, 1982).

The common ground needed may include quite specialized information as well. Suppose for the same two men you utter (9), (10), or (11).

(9) That man is a real Falstaff.

(10) That man is a real Don Quixote.

(11) That man is a real George Smith.

Your companion must appeal to the knowl-

edge he assumes the two of you share, as educated adults, of Falstaff and Don Quixote for (9) and (10), and as friends of thin George Smith for (11). An educated but unacquainted overhearer might guess (9) and (10), but could not interpret (11).

These arguments have been based on intuitions about demonstrative references in context. Experiment 4 was a field experiment designed to substantiate even subtler intuitions about the use of presuppositions.

#### EXPERIMENT 4

##### Method

For this experiment, we used a picture from *Newsweek Magazine* of President Ronald Reagan sitting with David Stockman, then the director of the Office of Management and Budget. We supposed that virtually everyone at that time would assume Reagan was very familiar to most people and Stockman less familiar. As interviewer, Buttrick approached 30 students one at a time on the Stanford University campus, showed them the picture, and, while recording their responses on a clipboard, asked half of them Question 1 and the other half Question 2.

Question 1. You know who this man is, don't you?

Question 2. Do you have any idea at all who this man is?

With Question 1, Buttrick appeared to presuppose the student would surely know who "this man" is. With Question 2, he appeared to presuppose the opposite. If people rely on such presuppositions, the students should tend to select Reagan for Question 1 and Stockman

for Question 2. To check our assumptions about Reagan's and Stockman's familiarity, Buttrick then asked each student to rate how recognizable Reagan and Stockman would each be to the general public on a 1 to 7 scale with 1 being "very low" and 7 being "very high." He also asked them to identify either man they had not yet named.

##### Results and Discussion

Questions 1 and 2 led to different choices of referents. The 30 responses are summarized in Table 4. For Question 1, 14 of 15 students chose Reagan, 12 outright and 2 tentatively; none chose Stockman. For Question 2, 7 of 15 students chose Stockman, 3 outright and 4 tentatively; only 2 chose Reagan and they did so tentatively. This difference is highly significant,  $\chi^2(1) = 15.65, p < .001$ . As expected, Reagan was judged more recognizable than Stockman, 6.2 to 2.3,  $t(29) = 30, p < .01$ . All 30 students identified both Reagan and Stockman correctly.

The perceptual model cannot account for these findings. If it were correct, there should be no difference between Questions 1 and 2, since the referent should be chosen entirely on the basis of perceptual salience. But there was a difference. By the common ground model, the students should use Buttrick's apparent presuppositions in their rationales for selecting Reagan versus Stockman. When he appeared to presuppose they knew the referent, they should select Reagan, reasoning that Reagan is better known than Stockman; when he appeared to presuppose the opposite, they should select Stockman. And that is what occurred.

TABLE 4  
RESPONSES TO QUESTION 1 ("YOU KNOW WHO THIS MAN IS, DON'T YOU?") AND QUESTION 2 ("DO YOU HAVE ANY IDEA AT ALL WHO THIS MAN IS?") IN EXPERIMENT 4

| Response                           | Question 1 | Question 2 |
|------------------------------------|------------|------------|
| 1. Selects Reagan                  | 12         | 0          |
| 2. Points at Reagan, "This one?"   | 2          | 2          |
| 3. Selects Stockman                | 0          | 3          |
| 4. Points at Stockman, "This one?" | 0          | 4          |
| 5. "Which one?"                    | 1          | 5          |
| 6. Identifies Reagan and Stockman  | 0          | 1          |
| Totals                             | 15         | 15         |

The students were more certain in answering Question 1 than Question 2. Let us call a response uncertain if it was a request for confirmation or clarification. Only 3 of 15 students were uncertain for Question 1, as compared with 11 of 14 students for Question 2,  $\chi^2(1) = 7.74$ ,  $p < .01$ . There are several possible reasons why. The students might have been more certain of their answer to Question 1 than to Question 2. Or they might have been more ready to assume Buttrick was referring to the more recognizable man, namely Reagan. Or Reagan might have been slightly more prominent perceptually. Whatever the reason, it does not affect the conclusion that the students relied heavily on the speaker's apparent presuppositions.

#### GENERAL DISCUSSION

Demonstrative references are not simple. To understand the reference in "How would you describe the color of this flower?" the students in Experiment 1 had to appeal to the "relevant context" and draw certain inferences. The "relevant context," we have argued, consists of the common ground between the speaker and addressees, and the inferences needed are based on the principle of optimal design, which governs language use in general (see Clark, 1983b). The students had to find a demonstrative relation  $F$  to get them from the demonstratum  $d$  (the picture of four flowers) to the referent  $r$  (one of the four flowers). The  $F$  could not be just any relation. It had to be the one they believed they were intended to infer on the basis of common ground.

To infer  $F$  for underdetermined demonstrations, we have argued, people try to select the object that both fits the descriptor and is the most salient against common ground. They do this in much the same way they find the best solution in a Schelling task. They assume the speaker and addressees "must 'mutually recognize' some unique signal that coordinates their expectations of each other" (as Schelling put it), and they take that signal to be the most salient object in their common ground. In Experiments 2 and 3, choices in the ref-

erence task were well accounted for by the corresponding Schelling choices.

People must weigh every part of common ground that might be pertinent. We have demonstrated the influence of four such parts:

1. *Perceptual salience*. Each demonstratum in our experiments consisted of an arrangement of objects, and students could estimate how people in general would view these arrangements. When there was no other pertinent information in common ground, they would simply select from the display the perceptually most salient object, which is the only object they could be expected to pick out uniquely. In Experiment 2, as predicted, the reference choices were well accounted for by the salience judgments.

2. *Speaker's goals*. When the speaker's goals are part of common ground, they are often crucial to the interpretation of an utterance (see Clark, 1978, 1979; Cohen & Perreault, 1979; Gibbs, 1981). The utterance "City Hall, please" would be interpreted one way by a taxi driver, but another way by a telephone operator because of what they presume to be common ground about the speaker's goals. The speaker's presumed goals are just as important for demonstrative reference. In Experiment 3, when it was common ground that the speaker was looking for a present for Uncle George, the listener's choice of referent was based on this goal in combination with perceptual salience.

3. *Speaker's assertions*. What the speaker asserts also becomes part of common ground, and it too can influence the listener's choice of referent. To interpret "That man weighs too much for his own good," the addressee has to use what was asserted about the referent. An overhearer unable to understand the assertion, as in "That is what George will look like soon," might be unable to pick out the right referent.

4. *Speaker's presuppositions*. What the speaker presupposes in his utterance is also part of common ground and potentially relevant. This was demonstrated in Experiment 4 for two kinds of presuppositions. When Buttrick said "You know who this man is, don't

you?" he appeared to presuppose explicitly that "this man" was a man the student would surely know. He also appeared to presuppose, implicitly, that it was common ground that Reagan was better known than Stockman. So for this question, the students readily chose Reagan, but for "Do you have any idea at all who this man is?" they tended to choose Stockman instead.

When people are addressed by someone, they ordinarily assume the speaker has done his best to enable them to understand him. It is on this assumption, the principle of optimal design, that they can reason through to the referent. Yet people also recognize that the speaker can make misjudgments—especially about common ground. When they detect misjudgments, they have several options. They can guess, with the possibility of revising their guess later; they can guess but ask for confirmation; or they can directly ask for clarification. In Experiments 1 and 4, students were indeed more likely to ask for confirmation or clarification the less well designed the reference was.

Demonstrative reference is perhaps the prototype of expressions that cannot be understood without appeal to context. But what context? If our proposal is correct, all the information the listener should ever appeal to is the speaker's and addressees' common ground. Similar claims have been made for the use and understanding of conventional and novel words, assertions, presuppositions, direct and indirect speech acts, and definite reference (see Clark, 1983a; Clark & Carlson, 1981; Clark & Marshall, 1981; for reviews). If these claims are correct, it will be crucial to study how people create, represent, and access common ground.

#### REFERENCES

- BENNETT, M. Demonstratives and indexicals in Montague grammar. *Synthese*, 1978, **39**, 1–80.
- CLARK, H. H. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 1973, **12**, 335–359.
- CLARK, H. H. Inferring what is meant. In W. J. M. Levelt & G. B. Flores d'Arcais (Eds.), *Studies in the perception of language*. London: Wiley, 1978.
- CLARK, H. H. Responding to indirect speech acts. *Cognitive Psychology*, 1979, **11**, 430–477.
- CLARK, H. H. Making sense of nonce sense. In G. B. Flores d'Arcais and R. Jarvella (Eds.), *The process of language understanding*. New York: Wiley, 1983a.
- CLARK, H. H. Language use and language users. In G. Lindzey and E. Aronson (Eds.), *Handbook of social psychology*. Reading, Mass.: Addison-Wesley, 1983b, Third edition.
- CLARK, H. H., & CARLSON, T. B. Context for comprehension. In J. Long and A. D. Baddeley (Eds.), *Attention and performance IX*. Hillsdale, N.J.: Erlbaum, 1981.
- CLARK, H. H., & CARLSON, T. B. Hearers and speech acts. *Language*, 1982, **58**, 332–373.
- CLARK, H. H., & MARSHALL, C. R. Definite reference and mutual knowledge. In A. K. Joshi, B. L. Weber, & I. A. Sag (Eds.), *Elements of discourse understanding*. Cambridge: Cambridge Univ. Press, 1981.
- COHEN, P. R., & PERRAULT, C. R. Elements of a plan-based theory of speech acts. *Cognitive Science*, 1979, **3**, 197–212.
- FILLMORE, C. J. Towards a descriptive framework for spatial deixis. In R. J. Jarvella and W. Klein (Eds.), *Speech, place, and action*. New York: Wiley, 1982.
- GIBBS, R. W., JR. Your wish is my command: Convention and context in interpreting indirect requests. *Journal of Verbal Learning and Verbal Behavior*, 1981, **20**, 431–444.
- HAWKINS, J. A. *Definiteness and indefiniteness: A study in reference and grammaticality prediction*. London: Croom Helm, 1978.
- LAKOFF, R. Remarks on *this* and *that*. *Papers from the Chicago Linguistic Society*, 1974, **10**, 345–356.
- LYONS, J. Deixis as the source of reference. In E. L. Keenan (Ed.), *Formal semantics of natural language*. Cambridge: Cambridge Univ. Press, 1975.
- LYONS, J. *Semantics, Vol. 2*. London: Cambridge Univ. Press, 1977.
- MACLARAN, R. On two asymmetrical uses of the demonstrative determiners in English. *Linguistics*, 1980, **18**, 803–820.
- NUNBERG, G. The non-uniqueness of semantic solutions: Polysemy. *Linguistics and Philosophy*, 1979, **3**, 143–184.
- OLSON, D. R. Language and thought: Aspects of a cognitive theory of semantics. *Psychological Review*, 1970, **77**, 257–273.
- PECHMANN, T., & DEUTSCH, W. The development of verbal and nonverbal devices for reference. *Journal of Experimental Child Psychology*, 1982, **34**, 330–341.
- SCHELLING, T. C. *The strategy of conflict*. Cambridge, Mass.: Harvard Univ. Press, 1960.

#### REFERENCE NOTE

- CLARK, H. H., & SCHREUDER, R. On demonstrative reference. Unpublished manuscript. Department of Psychology, Stanford University, Stanford, Calif., September 1981.