Controlling the intelligibility of referring expressions in dialogue

Ellen Gurman Bard
University of Edinburgh, Edinburgh, U.K.

Anne H. Anderson
University of Glasgow, Glasgow, U.K.

Catherine Sotillo
University of Edinburgh, Edinburgh, U.K.

Matthew Aylett
University of Edinburgh, Edinburgh, U.K.

Gwyneth Doherty-Sneddon
University of Glasgow, Glasgow, U.K.

Alison Newlands
University of Glasgow, Glasgow, U.K.

Accepted for publication in *Journal of Memory and Language* as of 21 May, 1999

Address for correspondence: Dr. E. G. Bard, Department of Linguistics, Adam Ferguson Building, University of Edinburgh, Edinburgh EH8 9LL, U.K. Email: ellen@ling.ed.ac.uk.

Abstract

If speakers articulate clearly enough to meet the perceptual needs of their listeners, clarity should depend on what listeners know about (listener-Given) rather than on what speakers know about (speaker-Given). For words excerpted from spontaneous speech, however, intelligibility to naive adult listeners showed only effects of the speaker's knowledge. Words introducing labeled map landmarks to two successive listeners were less clear on repetition even though the second listener had not heard the original mention (Experiment 1). Repeated mentions became less clear even after the listener reported inability to see the landmark (Experiment 2). Speakers were affected by what they had heard listeners mention: Intelligibility fell equally in co-referential repetitions across and within speakers (Experiment 3), whether or not the repeater could see the referent (Experiment 4). The results are explained via fast priming processes dependent on the speaker's knowledge and slow, optional processes drawing inferences about the listener's.

Keywords: intelligibility, referring expression, dialogue, spontaneous speech, hearer-Given, speaker-Given

Controlling the Intelligibility of Referring Expressions in Dialogue

No two spoken instances of a word are identical, even if they are produced by the same speaker in the same conversation. Yet variability in pronunciation is by no means random. Among the many controlling factors is the availability of information beyond the acoustic substance of the word token itself which might help a listener to decipher the speech sounds as the correct word. For this reason, it is often thought that the psychological processes involved in speech production must include the speaker's model of what the listener knows and perceives (Bolinger, 1963, 1981; Chafe, 1974; Lindblom, 1990). As Lindblom (p. 405) put it, "the speaker estimates the running contribution that signal-complementary processes will make during the course of an utterance" and continuously adjusts articulatory exactness to accommodate listeners' residual dependence on the speech signal. Lindblom suggested that in so far as other sources of information are available, speakers `hypo-articulate', but wherever the speech sound is the sole basis for recognition, they `hyper-articulate'. Since we will be discussing changes in pronunciation in terms of their effects on a listener's ability to recover the word from its spoken form, rather than the phonetic or phonological processes which achieve this result, we will use the term *intelligibility* for the variable that is controlled by hypo- and hyper-articulation.

This paper will test the proposal that adjustments to intelligibility are based on a model of the listener's, or more precisely, of the addressee's knowledge. Within this introduction, we will first exemplify some of the adjustments to word intelligibility which are thought to represent accommodation to the listener. We will show, however, that the existing evidence links them only to the speaker's own knowledge of the linguistic context, not to the listener's. Then we will explain why modeling listeners' knowledge is a very demanding task and present the current views about how speakers manage the modeling problem. Finally, we will explain how our experiments test for the effects of aspects of listeners' knowledge that speakers should be readily able to notice.

Changes in word intelligibility with context

The best-known examples of contextual effects on intelligibility depend on the relationship between a lexical item and the sentence context in which it is uttered. Lieberman (1963) showed that tokens of words like *nine* produced in (1) below were longer, had higher peak fundamental frequencies, and were more *intelligible*, in the operational sense of being more recognizable when excerpted from context, than tokens produced in contexts like (2).

1. The word you are about to hear is *nine.*
2. A stitch in time saves *nine*.

Intelligibility seemed to respond to redundancy: normal adults could easily supply the word *nine* given the remainder of (2), but not given the remainder of (1). Reliable negative correlations between redundancy and intelligibility have also been reported in more natural materials (Bard & Anderson, 1983; Fisher & Tokura, 1995; Fowler & Housum, 1987; Hunnicutt, 1985; Samuel & Troicki, 1998). This kind of relationship also seems to account for the careful articulation of words read in lists, where there are no contextual clues to word identity, and the much less careful delivery in running speech. While citation forms tend to be fully recognizable, Pollack and Pickett (1963) reported that, on average, only 50% of listeners could identify individual words isolated from surreptitiously recorded conversation.

A second kind of evidence, clearly akin to the first, will provide the focus of the current paper. This work shows that discourse status affects the way a word is articulated. Fowler and Housum (1987) described this effect as part of speakers' ability to signal the distinction between 'New' and 'Old' or 'Given' information. In the simple form of the Given/New distinction, New information is information just being introduced. Given information has been primed, or readied for use in discourse (Prince, 1981) by being made prominent in some way, whether by obvious physical presence at the time of speaking ('situationally evoked' in Prince's account), by previous mention ('textually evoked'), or by implication or association from information which has been explicitly mentioned ('inferrable'). Fowler and Housum's claim is that words which mention Given information can be interpreted in the context of that information, and as a consequence, can be articulated less clearly.

Though Given information often has been mentioned before, effects on intelligibility are associated with Givenness per se, rather than with the mechanics of repeated articulation. In Table 1, the findings which characterize the effect are summarized in terms of reference, Given status, and how that status is achieved. Since Given status is a characteristic of extra-linguistic information, a word token must refer to such information to induce the effect. Hence words merely read repeatedly in lists show no reliable order effects on intelligibility (Fowler, 1988), though two tokens of a word in a meaningful text do differ, with the second token, the one that refers to a Given entity, generally shorter and less intelligible when isolated than the first (Fowler, 1988; Fowler & Housum, 1987; Hawkins & Warren, 1994). Repeated referential use itself does not suffice. A relatively degraded word token is produced when the word refers to an entity Given by virtue of being mentioned previously (Fowler & Housum 1987), but not when the second use of the word refers to a New item of the same sort (Bard, Lowe, & Altmann, 1989). Nor is repetition a necessary condition. Introductory mentions of items Given by virtue of physical presence, (situationally Given in Prince's 1981 system) are less intelligible than introductory mentions which are truly New to the context (Bard & Anderson, 1994). Thus, Givenness without previous mention degrades intelligibility but repeated use without Givenness does not.

(Insert Table 1 about here)

These effects appear to parallel the well-documented influence of discourse status on the lexical and syntactic form of referring expressions (Ariel, 1990; Chafe, 1974; Clark & Haviland, 1977; Gundel, Hedberg, & Zacharski, 1993; Prince, 1981, 1992). The two phenomena also appear to be sensitive to the structure of extended discourse (Fowler & Levy, 1991; Fowler, Levy, & Brown, 1997; Vonk, Hustinx, & Simmons, 1992;). It is not surprising that the same explanation has been offered for the changes in intelligibility and the changes in referring expression: the needs of the listener. Prince (1992), for example, referred to Given status as “Hearer-old”. Ariel described a delicate adjustment of forms of referring expression to the antecedent's accessibility to the reader or listener. Gundel et al. described a Gricean Maxim of Quantity (Grice, 1975), which guides speakers and writers in choosing a form elaborate enough to meet the addressee's current needs. The effects of context on articulation might be ascribed to an analogous Maxim of Articulatory Quantity.

In all these cases -- the intelligibility-redundancy trade-off, the Givenness effect on intelligibility, and the effects of accessibility on form of referring expression -- the theory implicates speakers' models of listeners’ knowledge. As Table 1 indicates, however, the experiments on articulatory clarity manipulated only the speaker’s knowledge: The effect was found for mentions of entities which the speaker had previously mentioned, had heard mentioned, or could see at the time. Neither the listener’s knowledge nor the speaker’s observation of the listener was examined. Instead two assumptions were made: first, that even if speakers and listeners have different knowledge, Gricean maxims are accurately observed, and, second, as a consequence, that markers of Given status from a speaker will be reflexes of listener-Givenness.

This pair of assumptions attributes to speakers and writers a remarkable feat of maintaining continuous and accurate models of their addressees' knowledge. Yet this interpretation of the 'Gricean burden' is far too demanding to be taken for granted. More strictly construed, the Gricean maxims prescribe only assumptions that listeners should have about speakers’ behavior if they are to interpret material which does not initially seem to be suitable in quantity, quality or the like. Though they provide useful starting points for interpretation, the maxims cannot guarantee that the speaker can actually read the listener’s mind. Nor can they ease the considerable difficulty that a speaker of normal abilities would have in keeping track of and making proper inferences from all the evidence revealing the listener's pertinent knowledge.

Defaulting under the Gricean Burden

Maintaining an accurate model of the interlocutor is a potentially insuperable problem. As Clark and Marshall (1981; see also Keysar, 1997; Stalnaker, 1978) explained, speakers need to maintain an internal account of mutual knowledge, the collection of information which both speaker and listener know and which each knows they both know. For both to know the same things is difficult enough, but for each to know that both know the same things and that they understand the same thing by what they know, and so forth, is a potentially open-ended task, for every confirmation of agreement would need to be checked to be sure it masked no deeper disagreement. To have conversations in real time, Clark and Marshall suggested, any speaker must default to an optimistic working assumption about shared knowledge, that is, to the assumption that their own knowledge is a good model for the listener's. The question is when this default takes place. Three answers to this question have been offered.

The strongest position seems to be implicit in the notion of tailoring intelligibility on-line, though we have been unable to find discussion of this issue in the literature on controlling articulation. To achieve genuine adjustment to listeners' needs, speakers must seldom default to their own view, and only after strenuous efforts to model listeners as veridically as possible. To support continuous adjustment of articulation, speakers should observe listeners continuously for signs of misunderstanding or disagreement. Wherever speaker’s and listener’s knowledge differ, listener’s knowledge should take precedence. We might call this the *No Default Hypothesis*.

The second alternative is that speakers take special note of middle- or long-term characteristics of their listeners which affect likely overlap with their own knowledge. Various kinds of ‘co-presence’ in social or regional background (Fussell & Krauss, 1992; Isaacs & Clark, 1987), physical location during the interaction (Schober, 1993), or recent experiences (Brennan & Clark, 1996; Schober & Clark, 1989; Wilkes-Gibbs & Clark, 1992) are taken into account in adjusting the default assumption. Although this work is usually interpreted as showing that the ‘initial design’ (Horton & Keysar, 1996) of conversational speech is sensitive to the listener's needs, it does not directly address on-line processes throughout speech production. The original purpose of assessing co-presence (Clark & Marshall, 1981) was to provide speakers with reasonable grounds for defaulting to their own knowledge as a proxy for the more elusive facts about the listener's. Applied to the production of speech, this approach suggests that speakers should attend to evidence for and against co-presence and that defaulting should go on for some undefined time after positive evidence, without continuous recalculation. We might call this the *Co-presence Default Hypothesis*.

The third proposal makes a modular division between the initial formulation of utterances, a process based on the speaker's knowledge, and the monitoring and revision of output, processes based on a model of the listener's knowledge, or more precisely, of common ground. Called the *Monitoring and Adjustment Hypothesis* by Horton and Keysar (1996), this model defaults first and pays later. It offers the advantages of economy and speed. Faultless utterances, those for which speaker's and listener's knowledge are alike, are produced faster than they would be if accurate listener modeling were a pre-requisite. Poorly designed utterances can be revised in response to explicit requests from the listener. Those explicit requests achieve re-alignment between interlocutors without the need for continuous modeling of the listener's knowledge. For this reason, post-feedback utterances should reflect any aspects of listener knowledge which the feedback has added to the speaker's knowledge. Otherwise, the listener's knowledge should be irrelevant to production.

To decide which of these hypotheses, if any, accounts for speakers' behavior, we must first examine situations where the listener's knowledge is not confounded with the speaker's. To do this, we will draw words from running spontaneous speech produced while pairs of normal adults communicated a route best defined in terms of labeled landmarks on schematic maps of imaginary locations. The map design controlled speaker's and listener's knowledge independently. In four experiments, we will ask whether the Givenness effects in Table 1 are sensitive to what listeners have heard (Experiment 1), what they can see (Experiment 2), or what they have mentioned (Experiments 3 and 4). By examining word intelligibility and duration, we will try to discover whether speakers adjust to what the listener plainly does and does not know. The pattern of results will help us to re-assess the hypotheses outlined above.

Experiment 1

We begin by examining how speakers respond to disparities between their own experience of what has been said and the listener's. We use introductory mentions of the names of landmarks produced by Instruction Givers, the participants who were relating the route to their partners. Because some landmarks appeared only on one participant's map, neither player could be sure that they saw what their partner saw. We compare those introductory mentions with introductory mentions produced when a speaker who had already led one partner through a particular map began to lead a second partner through the same map. From the point of view of the Instruction Giver, the discourse status of any landmark mentioned in Trial 1 changed between trials. On its introduction in the first trial, it was genuinely New to the discourse. On the initial mention in the second trial, it represented Given information. For the two successive Instruction Followers, however, the status of the landmark was the same at both introductory mentions. Because neither had worked with the map before or heard the Instruction Giver mention the landmark, the item was New for both Instruction Followers. Thus, the introductory mention on Trial 2 was speaker-Given and listener-New. If the speaker's knowledge controls articulation, second trial speaker-Given introductions should be less intelligible than first. If the listener's knowledge controls articulation, however, no loss of intelligibility should be observed between successive listener-New introductions.

Method

Corpus. All materials in this and later experiments were drawn from the 128 unscripted conversations of the HCRC Map Task Corpus (Anderson et al., 1991), in which pairs of speakers, designated the Instruction Giver or the Instruction Follower, collaborated to reproduce on one player's map a route printed on the other's. Neither speaker could see the other's map. Despite their designations, no restrictions were placed on what either speaker might say.

Because the maps portrayed imaginary places, all information relevant to the task appeared on the maps. Of the landmarks critical to the reproduction of the route, only half matched exactly between partners' maps, with alternate landmarks encountered on any route mismatching in name, number, or location. Speakers were warned in advance that their maps would not match exactly but they were not told how often to expect mismatches or what sorts of mismatches there were. Over the whole corpus, 16 different basic maps were used

The participants, all undergraduates at the University of Glasgow, were recruited as pairs of friends, with two such pairs producing each set of 8 dialogues. Half the dialogues in each set were between pairs of friends and half between individuals who had not met before the recording session began. Every speaker served as Instruction Follower on two different maps and as Instruction Giver twice for a single map, each time with a different Instruction Follower. All 'quads' of participants followed a balanced design which put all first trials with a particular map in the first 4 of the 8 dialogues, and all second trials in the second 4.

Eight groups of 4 speakers participated in the *face-screened* condition with a flimsy barrier preventing partners from seeing one another . Within this group, familiarity of speakers using a particular map was counter-balanced. Another 8 quads worked in the *face-visible* condition with no barrier, but otherwise exactly replicated the design used for the first 8. Thus, over the whole design, each map was used the same number of times with faces screened and faces visible and by familiar and unfamiliar pairs.

After participating in a series of dialogues, each speaker read a list of landmark names covering the maps just used. These *citation forms* provide control tokens of landmark names in a condition which does not give rise to pragmatic effects (Fowler, 1988)[1].

All materials were recorded on the same DAT (Sony DTC1000ES) using one Shure SM10A close-talking microphone and one DAT channel per participant. All recordings were made under the same studio conditions.

Stimuli. The stimuli were introductory tokens of the landmark names uttered by the Instruction Givers in 2 trials using the same map but differing in the identity of the Instruction Follower. All were taken from full literal mentions of the landmark labels. Thus, if the landmark was labeled "Site of Forest Fire", only a referring expression containing all 4 words could provide tokens for this study.

As in all the experiments described in this paper, disfluent items and those which suffered from cross-talk from the other speaker were excluded. Selection of materials followed the multiple criteria for the experiment until the maximal design-conformant size was reached for the cell with the sparsest representation in the corpus with the minimal amount of lexical overlap within the cell.

For all experiments, words were excerpted from digitally recorded materials and digitized at a sampling rate of 16 kHz for use with ESPS WAVES software. Word onsets and offsets were determined by examining spectrogram and time-amplitude waveform representations and listening to the results of excerptions. Cut points were set at 0-crossings and the segment boundary conventions of Laver et al. (1989) were used.

To avoid ceiling effects in intelligibility experiments, the speech file for each original word was multiplied, sample by sample, by a 16 kHz file of random noise (where all sample values were in the range 0.5 to 1.5) of the same length. In each resulting stimulus, the amplitude of each sample was related to that of the original speech and each had the same sign as the sampled data value it replaced. The noise-overlaid speech files were downloaded to Digital Audio Tape with an inter stimulus interval (ISI) of 8 seconds.

Design. Forty-eight item triples (Trial 1 first mention to first listener, Trial 2 first mention to second listener, citation form) were used. Items were restricted to those mentioned both by speakers working with faces visible and by those with faces screened, to conform to the design of another experiment (Anderson, Bard, Sotillo, Newlands, & Doherty-Sneddon, 1997) run in the same sessions.

Items were taken from the speech of 20 speakers in the face-visible condition and 18 in the face-screened condition. Because there was some lexical duplication among the 48 items, the materials were divided between 2 sets, each distributed by Latin square among 3 groups so that no group contained more than one token of a lexical type and all contained equal representation of each cell of the design. The stimuli were then randomized with items belonging to another experiment.

Participants and Procedure. Six groups of 9 listeners were paid to serve as participants. In this and all other experiments reported in this paper, participants were undergraduates at the University of Glasgow, where the corpus had been recorded. No listener had any known hearing loss. Stimuli were played over headphones in a sound-proofed room.

Participants were told that all stimuli were real spoken words and were instructed to write down the words they heard. Responses were typed directly into the computer. Four practice items preceded test materials. Participants were able to ask questions about the procedure before test materials were presented.

Results

One item was lost by experimenter error. The analysis was therefore based on 47 item triples. We report analyses of intelligibility and duration for the data summarized in Table 2.

Intelligibility. We report direct measures of intelligibility, that is, the proportion of trials on which a word is correctly identified, and *intelligibility loss*, the difference between the intelligibility of a word token excerpted from running speech and the intelligibility of the citation form of the same word uttered by the same speaker. Statistics on raw intelligibility and intelligibility loss measures are always comparable. We will regularly report only the intelligibility loss analyses, where the critical outcome will more often depend on a main effect than on a comparison within an interaction. Raw score analyses will be cited only where baseline intelligibility is at issue.

Analyses of Variance (ANOVAs) were performed using participants in intelligibility experiments (by participants), word triples (by items), and participants in the Map Task (by speakers) as cases. Both by-participants and by-items have halves (1, 2) of the experiment as a grouping variable and trial (first, second) and face condition (screened, visible) as repeated measures. Because of the distribution of speakers' contributions across halves of the experiment, the by-speakers ANOVA simply had speakers nested in face condition and crossed with trial.

An egocentric adjustment to the speaker's experience should show more intelligibility loss vis-à-vis citation forms on the second trial, when landmarks are speaker-Given, than on the first, where they are speaker-New. Table 2 shows that instead of a uniform effect of trial on intelligibility loss (all F-values < 1), there was an interaction between face condition and trial, F_1 (1, 52) = 16.44, $MSE = 0.0413$, $p < .0002$; F_2 (1, 45) = 12.35, $MSE = 0.0492$, $p < .001$; F_{speakers} (1, 36) = 6.21, $MSE = 0.0286$, $p < .0174$. Face-screened dialogues showed the egocentric pattern, with intelligibility loss significantly greater in Trial 2 introductory mentions (.18), when the term is listener-New but speaker-Given, than in Trial 1 introductions (.07), when it is New to both participants. Face-visible Trial 2 items were as unclear as the face-screened tokens of the same words (.12), but now Trial 1 introductions were significantly less clear (.23) than either the face-visible Trial 2 or the face-hidden Trial 1 tokens. (All 3 pair-wise comparisons at $p < .05$ or better in Newman-Keuls tests by participants and by items. Only the trial effect within face-screened dialogues is significant by speakers.)

(Insert Table 2 about here)

Duration. Millisecond durations of stimuli were normalized in order to allow comparisons between cells not containing the same lexical items or not produced by the same speakers. Normalization was carried out in a method which owes much to the findings of Campbell and Isard (1991). These authors described the different distributions of segment lengths for different phonetic segments as belonging to the same mathematical type but having different means and variances. They hypothesized that within linguistic domains like words or syllables all the items might be shortened or lengthened by the same proportion of their individual variance. Though this system is not a perfect predictor of word duration, it is better than simpler normalizations. The current method is a version of the system, which, in pretests, was no less accurate than the original in modeling the durations of the present materials. Two simplifying assumptions were made: that each distribution of segment length was Gaussian for log duration and that all the segments had the same mean and standard deviation. Using a machine-readable dictionary to determine the number of segments, number of syllables, and presence of stressed syllables in the citation phonetic transcription of a word, we generated potential log total duration for each stimulus word with each segment taking on the value corresponding to its mean log duration plus some common *z*-value, which we call *k*. The *k*-transformed duration effectively provides a measure of compression/expansion relative to the expected length of a word with the given number of phonetic segments, number of syllables and lexical stress.

Cell means for *k* appear in Table 2. ANOVAs followed the by-items and by-speakers designs for intelligibility loss. Although running speech forms were predictably shorter than citation forms in *k*-units, F_2 (1, 46) = 355.04, *MSE* = 0.2327, $p < .0001$; F_{speakers} (1, 36) = 224.14, *MSE* = 0.1389, $p < .0001$, there were no main effects or interactions of interest. The tendency for shorter second introductions was not significant by items or by speakers ($F_2 < 1$; $F_{\text{speakers}} < 1$).

Discussion

Experiment 1 was intended to determine whether speakers adjust clarity of articulation in the light of what their listeners have already heard. Results differed across the two face conditions. With no visual channel for communication, speakers reduced clarity on introducing an item which was listener-New but speaker-Given. For this group, behavior was controlled by the speaker's experience and not the listener's. With a visual channel, speakers produced speaker-Given second-trial introductory mentions which were comparable to the degraded speaker-Given mentions of the same words in face-screened dialogues. For this group, however, first-trial introductory mentions, New for both speaker and listener, were even less intelligible.

Why were initial introductions so unintelligible when speakers and listeners could see each other? Using the HCRC Corpus, Anderson et al. (1997) have shown that lowered intelligibility in initial mentions occurs when interlocutors can look at one another but are not actually doing so. When the listener is looking at the speaker (about one word in five or six), initial mentions rise in intelligibility. Such tend to occur at points where players who cannot see one another would ask for assurance that they or their listeners have understood instructions correctly (Doherty-Sneddon et al., 1997). Anderson et al. suggested that intelligibility rises at such points because speakers interpret listeners' gaze much as they would interpret the overt feedback. Otherwise, speakers seem to assume that they and their listeners are fully aligned in their views of the dialogue, and so reduce clarity accordingly. There are no indications that the speakers' optimism is warranted: map task performance -- as measured by deviations of the Follower's drawn route from Giver's pre-printed version - was no better with faces visible than with faces screened. All that appeared to differ was the speakers' expectations that they were being understood.

If Anderson et al. (1997) are correct, this cell of the design is not in conflict with the rest of the results. Although speakers who cannot see their interlocutors may initially be the more cautious about introducing New entities, the next introduction is made with less articulatory effort. The actual level of effort in second trial introductions seems to be independent of visual cues from the listener, for it does not differ significantly between face conditions. Second trial introductions, then, depend on the speaker's experience, not the listener's.

It is possible that speakers ignored listeners’ knowledge, or more correctly, listeners’ ignorance, because it was covert, the background rather than the text of a fairly complex dialogue task. Both Anderson et al. (1997) and the Monitor and Adjust Hypothesis suggest that listeners can affect speakers by providing overt indications of what they know. In the next three experiments, we examined such cases.

Experiment 2

This experiment examines the importance to the speaker of the listener’s ability to see the object under discussion. The Corpus design forced participants to be continuously concerned with this kind of perceptual co-presence. Not only were routes designed around landmarks rather than coordinates or distances, but half the landmarks critical to the route differed in some way between participants' maps. The frequency of accurate feedback about mismatching landmarks correlates with accurate reproduction of the route (Anderson & Boyle, 1994).

Accordingly, it is difficult to see how a cooperative speaker could manage without an account of items which are visually Given for the listener. To test for effects of this sort of listener modeling, we contrast two kinds of repeated mentions. In both, a single speaker repeatedly mentioned the name of an unshared landmark, a landmark appearing only on that speaker's map. In one case, just after that first mention, the speaker learned that the landmark was not shared between the maps, because the listener promptly and explicitly denied having the item. In the other case, the listener either failed to supply this feedback or supplied misleading information, and the original speaker had no reason to believe that there was a problem.

A speaker who takes note of feedback and adjusts to listener knowledge should mitigate the effect of repeated mention on intelligibility, restricting intelligibility loss when the listener has declared that the referent is inaccessible. On the other hand, should speakers depend on their own knowledge, intelligibility loss ought to be the same in both cases, for in both, the item was always visually and verbally Given for the speaker

Method

Materials and design. All stimuli were word tokens uttered as part of the names of unshared landmarks. All tokens occurred within full literal mentions of the landmark as labeled on the map where it appeared. The first and second running speech tokens of each name were produced by a single speaker within a single dialogue with no intervening mention of the landmark in any form. For each of the 120 first and second mentions, there was also a citation form read by the same speaker.

Of the 120 triplets, 60 belonged to each of two feedback conditions (denial, no-denial). Feedback was determined by the behavior of the other participant between first and second mentions. In the denial condition, listeners explicitly -- and correctly – denied the existence of the feature on their map; in the no-denial condition, there was no such feedback.

Example (3) illustrates a self-repetition by the Instruction Giver (G) of *pelicans* with an intervening a denial from the Instruction Follower (F). An example of a self-repetition with no denial is presented in Example (4).

3. G: Stop, um, beside the `s' of `Saxon'.

   F: Okay.

   G: And, have you got *pelicans*?

   F: **No**.

   G: No. Um, go down about three to four centimeters from the Saxon barn vertically downwards.

   F: So I'm above the rope bridge?

   G: Just a bit above the rope bridge, yeah.

F: Okay.

G: And then go underneath where I've got *pelicans* towards the left-hand side.

4. F: What about *the banana tree*? Does it /

G: Oh.

F: come before *the banana tree*?

Materials were selected from 21 different speakers in face-screened dialogues and from 23 in face-visible dialogues. Only 12 speakers contributed triples to both feedback conditions. Triplets of words were balanced for feedback within each face condition: 32 denial and 32 no-denial triples were selected from face-screened; 28 denial and 28 no-denial from face-visible dialogues. Of the 60 triplets per feedback condition, 8 were matched verbatim across feedback, and the rest were matched for word length and frequency as far as possible. As many items as possible (47 of the no-denial group, 48 of the denial) were selected from first trial dialogues. Within feedback conditions, some landmark names recurred (10 within denial and 12 within no-denial), but never within the speech of a single speaker. Because of the duplication of word types created in these ways, materials were divided into two halves, so that no lexical item appeared more than once in either and each half was divided into 3 groups to provide a Latin square in which no sub-group heard more than one token of any word. Preparation of materials was otherwise as described for Experiment 1.

Participants and Procedure. The participants were 54 Glasgow University students, 9 in each of the 6 groups. Other details of the methodology were as described for Experiment 1.

Results

Table 3 contains cell means for intelligibility and normalized durations.

Intelligibility. ANOVAs were performed on intelligibility loss. Participants were nested in halves of the design and crossed with mention (first, second) and feedback (no-denial, denial). Items were nested in halves and in feedback and crossed with mention. Speakers were crossed with mention but nested in feedback, because the design was not fully crossed for feedback by speakers. Because the speaker design was not reflected fully in the ANOVA, Speaker results are included only for interest.

Instead of a main effect of mention on intelligibility loss (all F-values < 1), there was an interaction between feedback and mention, F_1 (1, 52) = 21.96, MSE = 0.0156, $p < .0001$; F_2 (1, 116) = 5.30, MSE = 0.0718, $p = .0231$; F_{speaker} (1, 54) = 5.77, MSE = 0.0319, $p = .0198$, but not the interaction which would have indicated a cooperative consideration for the listener's predicament. A cooperative adjustment would have given less intelligibility loss for tokens following denial than following no-denial, but intelligibility loss for second mentions did not differ significantly with feedback (.16 when listeners apparently could see the named item; .21 when they apparently could not). Instead, there was an effect of feedback condition on the introductory mentions. First mentions which failed to elicit accurate feedback were significantly more degraded (no-denial loss from citation = .24) than first mentions which succeeded (denial loss from citation = .13) (Newman-Keuls test by participants at $p < .01$). In fact the no-denial introductory mentions were more degraded than the second mentions which followed them (Newman-Keuls by participants at $p < .05$). In contrast, where accurate feedback occurred, second tokens were more degraded than first (Newman-Keuls by participants at $p < .01$).

This effect was not an artifact of baseline citation measurements. ANOVAs on raw intelligibility scores, using 3 levels of mention (first, second, citation) but otherwise identical to intelligibility loss designs, revealed the same effects as the derived measure. The Feedback (2) x Token (3) interaction was significant by participants, F_1 (2, 104) = 11.76, MSE = 0.0154, $p < .0001$, and approached significance by items, F_2 (2, 232) = 2.80, MSE = 0.0720, $p = .063$. Neither second mentions nor citation forms differed across feedback conditions by Newman-Keuls tests. The denial materials displayed the usual reduction in clarity from first to second mention ($p < .01$). The source of the interaction was what it appeared to be in the analyses of intelligibility loss, the unusual unintelligibility of those initial mentions not followed by accurate feedback. They were harder to identify than any other running speech tokens ($p < .01$).

(Insert Table 3 about here)

Duration. ANOVAs for normalized duration measures followed the by-items and by-speakers designs used for intelligibility. Unlike intelligibility, duration loss vis-à-vis the citation form changed significantly over successive tokens (0.568 for first tokens and 0.675 for second), F_2 (1, 116) = 6.77, MSE = 0.1001, $p = .0105$; $F_{speaker}$ (1, 54) = 7.28, *MSE* = 0.0683, $p = .0093$. There were no signs of curtailing reduction in the denial condition: the interaction between mention and feedback was not significant (0.519 and 0.658 for mentions preceding and following a denial; 0.617 and 0.691 for pairs with no denial intervening; $F_2 < 1$; $F_{speaker} < 1$).

Discussion

Far from mitigating the repetition effect in response to the listener's overt feedback, speakers continued to reduce intelligibility. Second mentions were equally degraded in intelligibility and duration regardless of listeners' responses to the initial mention. There is nothing in these results to indicate speakers' sensitivity to feedback about what the listener can see.

The results were complicated by an unexpectedly extreme loss of clarity in those introductory mentions which failed to elicit accurate feedback. The effect on intelligibility was not, however, matched by an effect on duration. We have extensively re-examined the data, -- by form of clause, face visibility, and inherent characteristics of the particular samples used in this experiment, for example --, and have been unable to find any confounding variable which could explain the particular lack of care taken by speakers here. It is, of course, possible that the causal relationship runs in the other direction. Poorly articulated tokens may not have elicited good feedback because the addressee simply did not hear or recognize them. More subtly, the association may have something to do with how listeners interpret more and less clear tokens.

Fowler and Housum (1987) (see also Terken & Nooteboom, 1987) found that second, less intelligible mentions were correctly classed as later mentions in a forced choice ('old'/'new') task. More interestingly, in probed recall these less intelligible second tokens were more likely to prime the recognition of words related to the original mention. When our speakers happened to produce poor tokens as first mentions, they might have failed to signal to their listeners that these were New landmarks which needed to be located on their maps. In fact, the probed recall results suggest that poor tokens will instigate searches for antecedents in the listener's model of the dialogue. In these cases, the fruitless search may have distracted listeners from more pertinent tasks like providing a suitable reply to their partner.

Poorly designed introductory mentions aside, the results of Experiment 2 seem to indicate that adjustments of intelligibility with repetition are not sensitive to what listeners can see. Even overt feedback indicating a tactically important form of co-presence had no effect on speakers' control of clarity. Those speakers were either not monitoring for feedback or not adjusting their messages in response.

Experiment 3

We have so far asked whether speakers are sensitive to two of the means by which items become listener-Given, by being heard or seen. Now we turn to situations where items are listener-Given because the listener has mentioned them. To do this, we compare the effects of repeated mentions of two kinds. Same-speaker items were introduced and mentioned a second time by the same person. Between-speaker items were introduced by one player and repeated by the other. If repeated mentions are degraded relative to introductory mentions no matter who provided those original tokens, we might have evidence that both listener mention and speaker mention can confer Given status.

Why should we expect speakers to take note of what their interlocutors mention while ignoring their feedback? First, the listener played a role close to the speaker's own in constructing the discourse, and speakers may be more sensitive to such contributions than to commentary. Second, what the other speaker said might be less critical in itself than in what it implied. In dialogues where the participants were encountering a map for the first time, there were no grounds for introducing an item other than spotting it on the map. Although Experiment 2 indicated insensitivity to what listeners saw, the case here is somewhat different. Experiment 2 contrasted overt feedback about what the listener could not see with overt or covert indications of shared visual resource. In the present experiment, introductory mentions are positive evidence, overt indications of shared visual resource. Finally, if speakers do not add the terms introduced by their interlocutors to a common set of Given entities, it is difficult to see how any approximation to common ground can be composed.

On the other hand, since our results so far show no sensitivity to fairly obvious aspects of listener knowledge, it may be that speakers give a special status to their own discourse. If this is the case, then the effect of repetition on intelligibility loss will be greater for same-speaker than for different-speaker examples.

<u>Method</u>

<u>Stimuli and Design</u>. All stimuli were word tokens uttered as part of the names of shared landmarks, that is, of landmarks identically located, depicted, and labeled on Giver's

and Follower's maps. All occurred in both the first and the second mention of the same landmark via the same referring expression within a single dialogue and with no intervening reference to the landmark in any other form. Of the repeated mentions, 48 were introduced and repeated by the same speaker, while the other 48 were introduced by one speaker and repeated by a different speaker. All items came from the speakers' first trial with a map: none was a second introduction in the sense of Experiment 1. In Example (5) below, *the extinct volcano* is repeatedly mentioned by the same speaker, while *the tribal settlement* is repeated by a different speaker.

5. G: Start at *the extinct volcano* and go down round *the tribal settlement*. And then

   F: Whereabouts is *the tribal settlement*?

   G: It's at the bottom. It's to the left of *the extinct volcano*.

Lexical identity was matched as far as possible: 18 lexical items appeared as both same-speaker and different-speaker repetitions, while the remaining word types were matched as closely as possible for syllable length and word frequency. Where feasible, speakers who produced both same-speaker and different-speaker tokens were selected in preference to speakers who contributed to a single cell. Materials were drawn from the speech of 12 participants in the face-visible condition and 14 in the face-screened condition.

For each of the word tokens uttered in running speech, a citation form produced by the same speaker was used as a control. Thus there were 4 stimuli for each experimental item: Running Speech Mention 1, the introductory mention from the dialogue; Citation 1, the citation form read by the speaker of Mention 1; Running Speech Mention 2, the second mention from the same dialogue; Citation 2, the citation form read by the speaker of Mention 2. For same-speaker repetitions, Citation 1 and Citation 2 are the same word token.

With 48 word types nested in each of 2 levels of repeater (same, different) and crossed with 2 of form (running speech, citation) and 2 of mention (first, second)), the design yielded a total of 384 stimuli. To avoid multiple encounters with the same lexical items, materials were divided into 2 groups and each group was distributed among 4 subgroups by Latin square with the same random order of lexical items across each Latin square.

Participants and procedure. Eight groups of 10 University of Glasgow undergraduates attempted to identify the words, with only one token of any lexical item heard by any one listener. Materials preparation, instructions, and response collection were as described for Experiment 1. Stimuli were presented over headphones in individual sound-proofed booths. After the introductory phases, the 48 stimuli were presented with no further break.

Results

Intelligibility. ANOVAs again used intelligibility loss relative to citation form as dependent variable. Participants were nested in halves of the experiment and crossed with mention and repeater. Items were nested in levels of repeater and halves and crossed with mention. Speakers were crossed with mention; but since not all of the 26 speakers appeared in both repeater conditions, speakers were treated as nested in repeater level, and by-speaker results are included only for interest.

(Insert Table 4 about here)

Table 4 shows mean intelligibility and *k*-normalized duration for all cells. Overall, repetition increased intelligibility loss relative to citation form (first mentions .15, second mentions .23). The main effect of repetition was significant by participants and approached significance by items, F_1 (1, 78) = 6.17, MSE = 0.0856, p = .0151; F_2 (1,92) = 2.97, MSE = 0.1069, p = .0884; $F_{speaker} < 1$. Repetition had the same effect for same-speaker and different-speaker repetitions (all token by repeater interactions, $F_1 < 1$)

Duration. ANOVAs followed the designs for by-item and by-speaker intelligibility analyses and showed the same pattern. Again overall reductions in duration relative to citation form were greater for second tokens (by k of 0.801) than for first (by 0.581), F_2 (1,92) = 14.22, MSE = 0.1638, p = 0.0003; $F_{speaker}$ (1, 28) = 7.50, MSE = 0.0909, p = 0.0106, and again there was no interaction between token and repeater condition, F_2 (1,94) = 2.67, n.s.; $F_{speaker} < 1$.

Discussion

The results of Experiment 3 suggest that both speakers can confer Given status on an entity by mentioning it, for repeated mentions were shorter and less clear than first mentions, regardless of who produced the first mention. Speakers did not seem to discriminate between their own introductory mentions and those of another speaker.

The design of Experiment 3 leaves the possibility, however, that the results were based on an egocentric consideration. All materials referred to landmarks appearing on the repeater's map. Since introductory mentions were restricted to items on the introducer's map, same-speaker words had to name shared landmarks which the repeater could see. To match landmark names across conditions, different-speaker items also had to be shared. The repetition effect in both conditions might be due to the fact that both items were situationally Given (Prince, 1981) for the repeater. The other player's introductory mention might have been irrelevant. Experiment 4 offers a control against this interpretation.

Experiment 4

To interpret the results of Experiment 3, then, we need to be sure that repeaters assigned Given status to items which they had heard mentioned but could not see. We cannot provide the necessary controls by applying the design of Experiment 3 to unshared landmarks, rather than shared, because speakers seldom, if ever, introduced landmarks they could not see. Instead we examine only different-speaker repetitions for the effects of what the repeater could see while uttering the second mention

Experiment 4 uses introductory and second mentions of the names of the landmarks which were either shared and visible to both speakers, or unshared and not visible to the repeater. If speakers assign Given status to what they can see but not to what their listeners have mentioned, then they should reduce intelligibility only when repeating the names of shared landmarks. This is the outcome which would reveal an artefact in Experiment 3. If speakers are sensitive to both listener-mention and visual context, they should reduce intelligibility when repeating the names of shared and unshared landmarks, but more so when mentioning a landmark which they can see. Finally, if listener mention is sufficient to convey Given status and visual context is unimportant, then reduction should be the same with both kinds of repetition.

Method

Materials and design. The materials were all spoken words excerpted from different-speaker repeated mentions which used the literal form of the landmark name and from the corresponding citation form from each speaker. For 48 such sets, the landmark referred to was unshared: it appeared only on the introducer's map. For another 48, the landmark was shared and appeared with identical form, location, and label on both speakers' maps. In the examples below, *the rope bridge* provides repeated mentions of a shared landmark, and *the machete* provides mentions of an unshared item.

6. F: Right. How far?

   G: Um, at the opposite side.

   F: To the opposite side. Is it underneath *the rope bridge* or to the left

   G: It's underneath *the rope bridge*. And then from the tribal settlement go straight up towards the rope bridge and over the rope bridge. Then down three steps and along to above the volcano.

F: Eh, d ... Is down three steps below or above *the machete*?

G: Ah. *The machete*'s not on my map.

Items were contributed by 20 speakers from face-visible dialogues and 13 from face-screened. All items were from first trial dialogues. There were 48 repeated mentions of shared landmark names, and 48 of unshared, with half of each group from face-visible dialogues and half from face-screened dialogues.

Because few landmark names appear both as shared and unshared features, it was not possible to match word types across sharedness conditions. Items were thus nested in levels of speaker's visual access (shared, unshared) and crossed with mention (first, second) and form (running speech, citation), giving 384 stimuli in all. Twelve of the 33 speakers whose repetitions were used contributed second mentions to both shared and unshared categories. Items were excerpted and distributed by Latin square into 4 groups. Methods for excerption and ISIs were as described for earlier experiments.

<u>Participants and Procedure</u>. Four groups of 9 participants each were drawn from the population sampled in the other experiments. Procedure remained the same.

<u>Results</u>

Table 5 shows raw means and differences from citation form.

<u>Intelligibility</u>. In analyses of intelligibility loss, participants were crossed with visual access and mention. Items were nested in visual access and crossed with mention. Because the design was not completely crossed by speakers, a by-speaker analysis, with speakers nested in levels of visual access and crossed with mention, was performed for interest only.

Overall, there was a robust effect of repetition, with intelligibility loss increasing from first (.19) to second (.36) mention, F_1 (1, 35) = 28.77, MSE = 0.0343, p < .0001; F_2 (1, 94) = 12.36, MSE = 0.1064, p = .0007; $F_{speaker}$ (1, 43) = 16.10, MSE = 0.0612, p = .0002.

There was no tendency whatever towards a more extreme repetition effect for shared landmark names (all Mention x Visual Access analyses, $F < 1$). A main effect for repeater's visual access, significant by participants and speakers, F_1 (1, 35) = 18.50, MSE = 0.0171, p = .0001; F_2 (1, 94) = 3.26, MSE = 0.1295, p = .0743, $F_{speaker}$ (1, 43) = 5.87, MSE = 0.0815, $p <$ .0197, showed both introducer and repeater degrading unshared item names (first mention, .23 and second, .41) more than shared (.15 and .30).

(Insert Table 5 about here)

Duration. The ANOVA for duration loss from citation form to running speech form (as measured in *k*) followed the design for intelligibility loss. Like intelligibility, duration showed an overall effect of repetition, with reduction increasing from first (0.601) to second (0.714) mentions, F_2 (1, 94) = 4.35, MSE = 0.1398, p = .0397; $F_{speaker}$ (1, 43) = 5.97, MSE = 0.0841, p = .0187).

The interaction between mention and visual access, though now more prominent, was significant only in the imperfect by-Speaker analysis, F_2 (1, 94) = 3.58, MSE = 0.1398, p = .0616; $F_{speaker}$ (1, 43) = 5.05, MSE = 0.0841, p = .0298. There was a tendency toward a larger effect of repetition on names of shared (first mentions by 0.538 *k*-units, second mentions by 0.753) than on names of unshared landmarks (first mentions by 0.664 *k*-units, second by 0.675). The main effect of shared visual access was now not significant ($F_2 < 1$; $F_{speaker}$ (1, 43) = 1.57, MSE = 0.21881, *n.s.*).

Discussion

In this experiment, cross-speaker repeated reference to both shared and unshared landmarks led to increased intelligibility loss and to reduction in length. For *k*-normalized duration the effect of mention was less uniform, with something close to an additional effect for visual access to what was being named. This pattern indicates that listeners may indeed confer Given status on items they introduce whether or not speakers can see the item. Visual access, if it works at all, works in addition to the effect of mention. Taken with the results of Experiment 3, however, the outcome shows similar effects on intelligibility of repeated mention by a single speaker or across speakers and with or without supporting visual context. What is treated as Given is what has been mentioned.

General Discussion

Table 6 summarizes the conditions under which we found an effect of Given status on intelligibility, including what speaker and listener have said, seen, or heard. Where a cell contains both a '+' and a '-', intelligibility loss occurred in both conditions. Experiment 1 showed that what the listener had heard, or rather had not heard, was unimportant to speakers who reduced clarity on introducing a landmark to a new listener. Instead, in treating second-trial introductions as Given, speakers responded to their own previous experience. Experiment 2 showed that feedback about what the listener could or could not see did not mitigate intelligibility loss in second mentions. Again, once the speaker had mentioned the item, it was treated as Given. On the other hand, Experiment 3 showed that what the listener mentioned was then treated as Given, with intelligibility loss virtually identical to repetitions of names which the repeater had introduced. Experiment 4 showed that the introductory mention by the listener sufficed to assign Given status, because intelligibility loss could occur even for names of objects which the repeater could not see.

All in all, the results seem to suggest that although listeners may introduce New information, no other manifestation of their knowledge (the right half of the table) affects the speaker's subsequent control of intelligibility. Table 6 indicates that we might state the results even more strongly. Only one column has the value '+' throughout. In all the cases we tested, the repeating speaker heard the original mention and then produced relatively degraded second mentions. If Given status is assigned to what the speaker has heard, then listener introductions conform to a general rule. Rather than supposing that speakers model one aspect of listener's behavior only, we can now assume that speakers can merely fail to note who introduced what into a dialogue.

(Insert Table 6 about here)

This conclusion appears more radical than any of the hypotheses characterized in the introduction to this paper. Some more radical stance seems to be warranted, however, for none of the hypotheses predicted all of the present results. The No Default Hypothesis predicted that speakers would continually monitor for evidence of listeners' knowledge and adjust their articulatory effort to such evidence in preference to any record of their own knowledge. Experiments 1 and 2 show that speakers may ignore important facts about listeners. Experiments 3 and 4 together seem to show equal status for what the speaker has introduced and what the listener has introduced, rather than a preference for adapting to listener knowledge. The Co-Presence Hypothesis allows for longer-term substitution of speaker knowledge for listener knowledge but only when evidence for co-presence triggers the default. Experiments 1, 2, and 4 also largely fail to meet these predictions, for they all show that differences in shared information – experience with a map or ability to see a particular landmark – failed to change speakers' behavior. The Monitor and Adjust Hypothesis predicted that dependence on speaker knowledge could be mitigated if the listener supplied feedback on misalignment between listener and speaker. Yet Experiment 2 reveals no mitigating effects of overt feedback

To explain the results, we turn instead to the model on which the Monitor and Adjust Hypothesis is based, a model presented in Brown and Dell (1987) and Dell and Brown (1991). This model goes further than the Monitor and Adjust system in assigning different kinds of information to different stages in the formulation of utterances. In an attempt to model speakers' descriptions of pictured activities performed with unusual instruments, Brown and Dell proposed that speakers begin by structuring their basic message around a constellation of information from their own personal scenarios. Only gradually, via iterative monitoring and embellishment, do they adjust their output to non-prototypical information or to the listener's needs. Consigning all early stages of proposition design exclusively to the speaker's knowledge, Brown and Dell suggest that the repetition effects of Fowler and Housum (1987) are due to intra-lexical priming of the kind found in the production of semantically related word pairs (Balota, Boland, & Shields, 1989). Listener models are not available at this stage.

The present results seem to conform to a variant of this model. Like Dell and Brown (1991), we propose that at least two distinct kinds of processes account for speakers' control of the intelligibility of words in referring expressions. The first are fast automatic priming processes, the second slower processes which may demand elaborate and computationally taxing inference. We add a further complication: More demanding tasks compete for time and attention with dialogue and task planning.

The fast priming processes operate during any attempts to produce spoken utterances and closely resemble well-established effects in perception. As in perception, priming here depends exclusively on what a single individual experiences. Though we cannot discount Dell and Brown's (1991) suggestion that lexical-lexical priming is involved, context effects on intelligibility seem to demand two more kinds of priming.

The first of these depends on discourse information. Since Given information in written texts is activated information (see McKoon & Ratcliff, 1980), all entities introduced into a dialogue can be activated, even if they are not fully grounded as mutual knowledge by explicit interactions between interlocutors. Activation of the representation of referent objects primes their names (Mitchell & Brown, 1988). Primed names, as in the Balota et al. (1989) study, are produced faster, and the usual result of fast speech is decreased articulatory detail.

The second kind of priming depends on immediate linguistic context. Sentence contexts prime the recognition of single words which continue them. (see for example O'Seaghdha, 1997) to the extent that those words are syntactically and semantically appropriate continuations of the context. Once again primed words should reduce in length and intelligibility. Redundancy, in the sense of Lieberman's (1963) study, translates into priming in this interpretation. Like the repeated mention effects, the effects of sentence context are strongest in spontaneous speech (Fowler, 1988; Pedlow & Wales, 1987; Samuel & Troicki, 1998), where sentence structure and word pronunciation are being planned simultaneously.

Because no inferences or decisions are involved, the priming processes are quick. Because Given status is the prime for the Givenness effect, it is unimportant who makes the introductory mention, so long as the speaker observes it. If articulatory adjustments show any consideration for the listener, it is only by coincidence. In fact, consideration could be simulated to the extent that speaker and listener share knowledge to begin with or to the extent that the listener populates the conversation with overt references to information which the two did not initially share.

The second major process, or family of processes maintain, elaborate, and exploit what is usually supposed to be the speaker's model of the listener. Such processes might include deciding which kinds and degrees of co-presence license which varieties of default. They might include updating memory for dialogue events. They might also include determining which beliefs and which goals must be attributed to interlocutors on the basis of what they say, and in particular on the basis of any feedback they provide to the speaker's own utterances. As Brown and Dell (1991) and Horton and Keysar (1996) and their colleagues suggest, these complex processes are simply too slow to precede every attempt at speech production. They are therefore likely to be too slow to make their effects felt on a word-to-word basis in running speech.

Whether these processes are ever completed will depend on at least their complexity, the time at the speaker's disposal (Horton & Keysar, 1996), and the other demands on the speaker's attention. Carefully pre-planned utterances in formal lectures, for example, may be tailored over many iterations to match the outcome of these processes. Written output of this kind has the advantage of a stationary target: The only change in the putative addressee is in his or her experience of the growing text. In theory, it is possible to allocate sufficient time to assure completion of the revision cycles. At the other extreme, speakers performing a complex task with an interlocutor who could take the floor at any pause may never complete the computation and may not even have the spare capacity to acquire the information on which the computation is based. Although genuine errors of expression and understanding may result, the risk appears to be warranted (for a discussion of risky referential strategies, see Carletta & Mellish, 1996).

This model seems to accommodate both the present results and others in the literature. In this study, it helps to explain why the on-line control of production is sensitive to what listeners mention, but not to what to they offer by way of feedback or co-presence: Mentions fuel priming but feedback needs to be interpreted.

The model also helps to explain why the visual channel is not so much used as abused. Although speakers took the opportunities presented by a visual channel to become more careless in articulating introductory mentions, they did not monitor the listener's face carefully for signs of comprehension or confusion. Speakers with open sight lines actually looked at their listeners during only about one word in eight (Anderson et al., 1997). A typical figure for dialogue, this outcome suggests that social restrictions on inter-speaker gaze (Argyle, 1990) and the detrimental effect of gaze on the process of planning discourse

(Beattie 1978, 1980, 1981; Exline & Winters, 1965) are more compelling than the need for fine-grained access to the listener's reactions.

The model also makes some sense of the distribution of results across tasks. In general, the easier the task, the more firmly listener variables are found to affect speakers' behavior. For example, Horton and Keysar's (1996) speakers had to enable listeners to decide whether a description was true of a single simple shape on each trial. With no time pressure, speakers preferred to describe the figures relative to shared context rather than unshared. With time pressure, this cooperative preference disappeared. Similarly, more cooperative results are reported for experiments on the tangram task (e.g., Wilkes-Gibbs & Clark, 1992) than on the map task. The tangram task requires matching from a fully shared closed set of figures to a simple array. The main problem for the instructor in this task is labeling the next of a shrinking set of candidate figures. The map task is more difficult. Since landmarks may have different names on the Giver's and Follower's maps, there are occasional problems of labeling. Each player holds unique critical information in the form of unshared landmarks and initially neither player knows which critical items the other may have or exactly how many remain in play as the game proceeds. The demands of aligning partial representations under these conditions should leave little time for computationally expensive cooperative behavior.

Finally, the model predicts trouble. Articulation geared to the speaker’s experience should sometimes be inappropriate to the listener's needs. Since articulatory clarity affects how listeners interpret the spoken referring expressions (Fowler & Housum, 1987; Terken & Nooteboom, 1987), these misalignments should have consequences. We seem to have come across one such consequence in Experiment 2. The introductory mentions were extraordinarily unclear in just those cases where listeners subsequently failed to report the important news that they lacked the landmark. The inappropriate first mentions may not have alerted them to scan for a new item.

We can see three major objections to this approach. First, it is not immediately clear why Given status should prime more than mere repetition. Second, our model does not appear to predict the copious evidence for orderly variation in form of referring expression. Third, if adequate communication is the goal that controls language use, it is unclear how homo loquens could develop such an egocentric mode of operation. In all three cases, however, the literature allows for interpretations consonant with the present model.

First, the particular status of Given entities in priming of reduced pronunciations seems out of step with the many cases in which mere experience of the prime affects latency in perception tasks (for reviews, see Balota, 1994; Zwitserlood, 1996). Although Balota et al. (1989) demonstrated priming of word duration in an experimental paradigm which also yielded priming of response latency, the two effects are not necessarily identical. Balota et al. used delayed naming of words presented with or cued by associated and semantically related words. To demonstrate duration priming, Shields and Balota (1991) used modified report of previously read sentences. The nature of their study makes the demonstration of repetition priming less compelling than the demonstration of associative priming. Shields and Balota measured the duration of a single word per sentence, like *cat* in (7) - (9) below (p. 49), which might be preceded by the same word (7), a related word (8) or an unrelated control (9). The repeated and related condition targets were both shorter than the control target.

7. Her cat chases our cat under the table.
8. Her dog chases our cat under the table.
9. Her son chases our cat under the table.

Because no target item was co-referential, it might be argued that this experiment did show simple repetition priming. The difficulty is that it is almost impossible to utter items like (7) without contrastive stress on the words preceding both tokens of *cat*, and consequent de-accenting of the target. Associated items like (8) and unrelated cases like (9) might more naturally include accented tokens of *cat*. The fact that only items like (7) differed from controls in amplitude, as would be expected with a change in accenting, bolsters the suggestion that priming and duration effects have been obscured by a global prosodic artifact. If only associative and semantic priming of duration are firmly demonstrated, we are not so much contradicting the findings for duration priming as extrapolating from them.

Second, the model is less in conflict with the literature on referring expressions than it at first seems. These studies all show that speakers and writers make adjustments. It is not completely clear that they make the adjustments which listeners and readers need. The studies of accessibility in texts, as we noted earlier, observe the relationship between anaphor and antecedent, but do not directly test either accessibility to readers, or writers' attempts to track readers' needs. Studies of spoken interactions, typically between an instruction giver and a sequence of followers, do offer contrasting cases of listener needs, and contrasting speaker behavior, but they do not show that the two are closely linked. Schober (1993), for example, found changes in ostensible viewpoint in speech to listeners with different physical orientations, but less adjustment when the listeners were present than when they were merely described. Wilkes-Gibbs and Clark (1992) found several differences in the way tangram directors approached new matchers differing in their experience of the directors' earlier trials with other matchers. Some differences of matcher experience induced differences in the time and words which directors devoted to the next trial, but we cannot tell whether these adjustments actually met the new matchers' needs. If they did, that is, if directors supplied what each new matcher lacked, then we might expect all new matchers, whatever their experience, to have to make equally sparse contributions to the interaction themselves. No matcher figures are given which would test this prediction. Directors also varied the distributions of their referring expressions across a four-way accessibility hierarchy. The results accord nicely with prediction, but there are two difficulties. First, without contextual support of the kind that would invite artifacts, a three-way critical distinction within this hierarchy represents one of the most difficult discriminations to make in recognizing, let alone transcribing, spontaneous speech. Two metrically weak, very short syllables with centralized, often devoiced nuclei, one with 'real' frication (definite *the man*), the other with 'nuisance' frication (indefinite *a man*), must be distinguished from one another and from something which is not a syllable but just background noise (bare nominal *man*). With much more easily distinguished alternatives (e.g., *big* and *small*), Horton and Keysar (1996) observed that time pressure made speakers indifferent to what listeners knew. Second, we do not know that the suitable referring expression has the Gricean effect predicted for it, easing the matcher's way. In Anderson and Boyle's (1992) study of introductory mentions to naïve map task listeners, however, the definite/indefinite choice had no effect whatever on any subsequent measures of appropriate feedback or accurate task completion. At best, then, the literature on referring expressions and listener modeling does not present a clear counter-argument to the model outlined above.

Finally there is the issue of how speech production could possibly be designed as poorly as the present model suggests, with adjustment to listeners' needs conditional on time and attention commensurate with a complex computational task. We suggest that the design is very likely to be optimal because the default case, the case in which the speaker proceeds from his or her own knowledge, is usually adequate. Most conversations either in individual lifetimes or in the history of the species are likely to involve co-present interlocutors who are close together in time, space, and linguistic and social community. Much of what the interlocutors bring with them to the conversation and much of the way they view its contents will be similar. There may be little pressure to develop faster computation in dialogue because we can do surprisingly well without it. Approximate accuracy of this kind is often the end-result of environmental pressures on perceptual systems. For example, oyster catchers survive as a species despite their tendency to retrieve the largest rounded object in their environs when their single egg rolls out of the nest (Tinbergen, 1951). Because the egg usually is the largest round object visible, `largest-round-object' is usually an adequate model of an egg. Our present results point to nothing more radical than a similarly approximate adjustment to the likely internal state of likely conversational partners. Adapting to actual but less likely internal states may be an evolutionary and cognitive luxury.

References

Anderson , A. H., & Boyle, E. A. (1994). Forms of introduction in dialogues: Their discourse contexts and communicative consequences. Language and Cognitive Processes, **9,** 101–122.

Anderson, A. H., Bader, M., Boyle, E., Bard, E. G., Doherty, G., Garrod, S., Isard, S. D., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H. S., & Weinert, R. (1991). The H.C.R.C. Map Task Corpus. Language and Speech, **34**, 351–366.

Anderson, A. H., Bard, E. G., Sotillo, C., Newlands, A., & Doherty-Sneddon, G. (1997). Limited visual control of the intelligibility of speech in face-to-face dialogue. Perception and Psychophysics, **59,** 580–592.

Argyle, M. (1990). Bodily communication. London: Routledge.

Ariel, M. (1990). Accessing noun-phrase antecedents. London: Routledge/Croom Helm.

Balota, D. A. (1994). Visual word recognition: The journey from features to meaning. In M. Gernsbacher (Ed.), Handbook of psycholinguistics. (pp. 303–358). San Diego: Academic Press.

Balota, D. A., Boland, J. E., & Shields, L. W. (1989). Priming in pronunciation: Beyond pattern-recognition and onset latency. Journal of Memory and Language, **28**, 14-36.

Bard, E. G., & Anderson, A. H. (1983). The unintelligibility of speech to children. Journal of Child Language, **10,** 265–292.

Bard, E. G., & Anderson, A. H. (1994). The unintelligibility of speech to children: Effects of referent availability. Journal of Child Language, **21,** 623–648.

Bard, E. G., Lowe, A., & Altmann, G. (1989). The effects of repetition on words in recorded dictations. In J. Tubach & J. Mariani (Eds.), Proceedings of Eurospeech '89: Vol. 2. Proceedings of EUROSPEECH '89 (pp. 573–576). Paris: European Speech Communication Association.

Bard, E. G., Sotillo, C., Anderson, A. H., Doherty-Sneddon, G., & Newlands, A. (1995). The control of intelligibility in running speech. In K. Elenius & P. Branderud (Eds.), Proceedings of the XIIIth International congress of Phonetic Sciences: Vol. 4. Proceedings of the XIIIth International Congress of Phonetic Sciences (pp. 188–191). Stockholm: Stockholm University.

Beattie, G. (1978). Sequential patterns of speech and gaze in dialogue. Semiotica, **23,** 29–52.

Beattie, G. W. (1980). The role of language production processes in the organization of behaviour in face-to-face interaction. In B. Butterworth (Ed.), Language Production: Vol 1, Speech and Talk (pp. 69–107). London: Academic Press.

Beattie, G. W. (1981). A further investigation of the cognitive interference hypothesis of gaze patterns in conversation. British Journal of Social Psychology, **20,** 243–248.

Bolinger, D. (1963). Length, vowel, juncture. Linguistics, **1,** 5–29.

Bolinger, D. (1981). Two kinds of vowels, two kinds of rhythm. Bloomington, Indiana: Indiana University Linguistics Club.

Brennan, S., & Clark, H. (1996). Conceptual pacts and lexical choice in conversation. Journal of Experimental Psychology: Learning, Memory, and Cognition, **22,** 1482–1493.

Brown, P., & Dell, G. (1987). Adapting production to comprehension -- the explicit mention of instruments. Cognitive Psychology, **19,** 441–472.

Campbell, W. N., & Isard, S. D. (1991). Segment durations in a syllable frame. Journal of Phonetics, **19,** 37–47.

Carletta, J., & Mellish, C. (1996). Risk-taking and recovery in task-oriented dialogue. Journal of Pragmatics, **26,** 71–107.

Chafe, W. (1974). Language and consciousness. Language, **50,** 111–133.

Clark, H. H., & Haviland, S. E. (1977). Comprehension and the given-new contract. In R. O. Freedle (Ed.), Discourse production and comprehension (pp. 1–40). Hillsdale, NJ: Erlbaum.

Clark, H. H., & Marshall, C. R. (1981). Definite reference and mutual knowledge. In A. K. Joshi, B. Webber, & I. Sag (Eds.), Elements of discourse understanding (pp. 111–222). Cambridge: Cambridge University Press.

Dell, G., & Brown, P. (1991). Mechanisms for listener-adaptation in language production: Limiting the role of the "model of the listener". In D. J. Napoli & J. A. Kegl (Eds.), Bridges between psychology and linguistics: A Swarthmore Festschrift for Lila Gleitman (pp. 111–222). Hillsdale, NJ: Erlbaum.

Doherty-Sneddon, G., Anderson, A. H., O'Malley, C., Langton, S., Garrod, S., & Bruce, V. (1997). Face-to-face and video-mediated communication: A comparison of dialogue structure and task performance. Journal of Experimental Psychology: Applied, **3,** 105–125.

Exline, R. V., & Winters, L. C. (1965). Effects of cognitive difficulty and cognitive style upon eye contact in interviews. Paper presented at the Eastern Psychological Association.

Fisher, C., & Tokura, H. (1995). The Given-New contract in speech to infants. Journal of Memory and Language, **34,** 287–310.

Fowler, C. (1988). Differential shortening of repeated content words produced in various communicative contexts. Language and Speech, **28,** 47–56.

Fowler, C., & Housum, J. (1987). Talkers' signalling of 'new' and 'old' words in speech and listeners' perception and use of the distinction. Journal of Memory and Language, **26,** 489–504.

Fowler, C., & Levy, E. (1991). Some ways in which forms arise from functions in linguistic communications. . In P. Romeas (Ed.), Actes du XIIIeme Congres International des Sciences Phonétiques: Vol. 1. (pp. 279–282). Aix-en-Provence: Université de Provence.

Fowler, C., Levy, E., & Brown, J. (1997). Reductions of spoken words in certain discourse contexts. Journal of Memory and Language, **37**, 24–40.

Fussell, S. R., and Krauss, R. M. (1992). Coordination of knowledge in communication: Effects of speakers' assumptions about what others know. Journal of Personality and Social Psychology, **62,** 378-391.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), Syntax and semantics: Vol. 3. Speech acts (pp. 41-58). New York: Academic Press.

Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. Language, **69,** 274–307.

Hawkins, S., & Warren, P. (1994). Phonetic influences on the intelligibility of conversational speech. Journal of Phonetics, **22,** 493–511.

Horton, W., & Keysar, B. (1996). When do speakers take into account common ground? Cognition, **59,** 91–117.

Hunnicutt, S. (1985). Intelligibility vs. redundancy - conditions of dependency. Language and Speech, **28,** 47–56.

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

Keysar, B. (1997). Unconfounding common ground. Discourse Processes, **24**, 253-270.

Laver, J. D., Hieronymus, J., Bennett, C., Cohen, I., Dalby, J., Davies, D., Hiller, S., McAllister, M., & Purves, E. (1989). The ATR/CSTR Speech Database Project (Report 3). Edinburgh: Centre for Speech Technology Research, Edinburgh University.

Lieberman, P. (1963). Some effects of the semantic and grammatical context on the production and perception of speech. Language and Speech, **6,** 172–175.

Lindblom, B. (1990). Explaining variation: A sketch of the H and H theory. In W. Hardcastle & A. Marchal (Eds.), Speech production and speech modelling (pp. 403–439). Dordrecht, Netherlands: Kluwer Academic Publishers.

McKoon, G., & Ratcliff, R. (1980). The comprehension processes and memory structures involved in anaphoric reference. Journal of Verbal Learning and Verbal Behavior, **19**, 668-682.

Mitchell, D. B., and Brown, A. S. (1988). Persistent repetition priming in picture naming and its dissociation from recognition memory. Journal of Experimental Psychology: Learning Memory and Cognition, **14**, 213-222.

O'Seaghdha, P. G. (1997). Conjoint and dissociable effects of syntactic and semantic context. Journal of Experimental Psychology: Learning, Memory, and Cognition, **23,** 807-828.

Pedlow, R., & Wales, R. (1987). Some effects of context on spoken language production. Language and Speech, **30**, 373-385.

Pollack, I., & Pickett, J. (1963). The intelligibility of excerpts from conversation. Language and Speech, **6,** 165–171.

Prince, E. (1992). The ZPG letter: Subjects, definiteness, and information-status. In W. Mann & S. Thompson (Eds.), Discourse description: Diverse linguistic analyses of a fund-raising text (pp. 295–326). Amsterdam: Benjamins.

Prince, E. (1981). Toward a taxonomy of given-new information. In P. Cole (Ed.), Radical pragmatics (pp. 223–255). New York: Academic Press.

Samuel, A. G., & Troicki, M. (1998). Articulation quality is inversely related to redundancy when children or adults have verbal control. Journal of Memory and Language, **39,** 175–194.

Schober, M. (1993). Spatial perspective taking in conversation. Cognition, **47,** 1–24.

Schober, M., & Clark, H. H. (1989). Understanding by addressees and overhearers. Cognitive Psychology, **21**, 211–232.

Shields, L.W., & Balota, D. A. (1991). Repetition and associative context effects in speech production. Language and Speech, **34,** 47-55.

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

Terken, J., & Nooteboom, S. (1987). Opposite effects of accentuation and deaccentuation on verification latencies for Given and New information. Language and Cognitive Processes, **2,** 145–163.

Tinbergen, N. (1951). The study of instinct. London: Oxford University Press.

Vonk, W., Hustinx, L., & Simmons, W. (1992). The use of referential expressions instructuring discourse. Language and Cognitive Processes, **7**, 301–333.

Wilkes-Gibbs, D., & Clark, H. H. (1992). Coordinating beliefs in conversation. Journal of Memory and Language, **31**, 183–194.

Zwitserlood, P. (1996). Form priming. Language and Cognitive Processes, **11**, 589-596.

Author Note

Ellen Gurman Bard, Catherine Sotillo, and Matthew Aylett, Human Communication Research Centre and Department of Linguistics; Anne H. Anderson, Alison Newlands (now at Strathclyde University, Glasgow, U.K.), and Gwyneth Doherty-Sneddon (now at the Department of Psychology, University of Stirling, Stirling, U.K.), Human Communication Research Centre and Department of Psychology.

This work was supported by the ESRC (UK)-funded Human Communication Research Centre. A preliminary report of this work was presented by Bard, Sotillo, Anderson, Doherty-Sneddon, and Newlands (1995).

Correspondence concerning this article should be addressed to Dr. E. G. Bard, Department of Linguistics, University of Edinburgh, Adam Ferguson Building, George Square, Edinburgh EH8 9LL, U.K. Electronic mail may be sent to ellen@ling.ed.ac.uk.

Footnote

[1] For other details of design see Anderson et al. (1991). Sample maps and all transcriptions can be viewed and materials used in the present experiments can be played via links given on http://www.ltg.ed.ac.uk/~amyi/maptask/intelligibility_materials.html.