Dialogue

8

Dialogue cannot exist without humility.

PAULO FREIRE

C: What about you personally?
D: I am a social person, so when I feel trapped and alone I become extremely sad or depressed.
C: And what kinds of things make you feel angry?
D: When someone hurts or disrespects me or someone I care about, I feel incredibly upset and angry.
C: But do they feel differently to you on the inside?
D: Yeah, they do. Happy, contentment, and joy feel more like a warm glow on the inside. Sadness, depression, anger, and stress feel much heavier and more weighted down.

One of the preceding conversational partners is a person, one is a computer program. Can you tell which is which? If you can't, the computer program has passed your personal version of the Turing test, named after the computer science and code-breaking genius, Alan Turing. In fact the computer program LaMDA produces such convincing responses, that some people involved in developing the technology believe that it has achieved not just an extremely sophisticated approximation of human language but also actual consciousness. (Spoiler alert: C is computer scientist Blake Lemoine. D is the artificial intelligence, or AI, program LaMDA.) We might ask therefore what LaMDA does to engage in dialogue and how those computations compare to what humans do. Let's start with the humans and return to AI at the end of the chapter.

Introduction to Psycholinguistics: Understanding Language Science, Second Edition.
Matthew J. Traxler.

Companion Website: www.wiley.com/go/traxler/psycholinguistics2e

Studying human dialogue is challenging because we have to take into account the joint and separate actions of two individuals, both of whom contribute to the conversation. Participants in conversation speak and listen, and an individual's comprehension and production processes overlap in time. Communication occurs via the words that participants speak, the *main channel*, but it also occurs via gestures and other forms of *backchannel* signals (such as head-nodding, *m-hmm* noises, and facial gestures) that listeners emit to indicate how well they are understanding what the speaker is saying (or at least, how well they think they are understanding the speaker). Because dialogue involves (at least) two participants, it helps to separate out the "speaker" and "listener" roles as we consider the factors that influence the behavior of speakers and listeners in a dialogue. But keep in mind that treating the speaker and listener roles as separate is really a matter of convenience.

This chapter reviews major theories of dialogue that seek to predict and explain how speakers and listeners cooperate to exchange thoughts, beliefs, and information in conversation. The chapter begins with an overview of Paul Grice's ideas about the rules that speakers follow when crafting utterances. Next, it takes up Herb Clark's ideas about how speakers and listeners cooperate to expand the amount of information that they hold in common. Third, it considers how much weight speakers give to listeners' needs while they are planning and producing utterances (whether speakers are *cooperative* or *egocentric*). It also considers whether listeners pay attention to speakers' knowledge when they interpret speakers' statements (or whether listeners, too, are egocentric).

Gricean Maxims

When people converse, they use words in combinations to express ideas. Whether a contribution in a conversation makes sense depends on more than just the meanings of the individual words and the literal meaning of the combination. For example, it makes perfect sense in isolation to say *Susan has nice hair*, and the normal force of that statement would be something like "the hair on Susan's head is attractive" or "most people would like to have hair like Susan's." But what if that statement were set in a particular context, say the context of a letter of recommendation? Let's say Susan is applying for a job and her former boss writes:

> Dear Potential Employer, I encourage you to hire Susan. Susan has nice hair.
> Sincerely, Susan's Former Boss.

Most likely, the reader will understand the propositional content of the statement ("Susan's hair = good"). But even though the reader can understand what the boss said, the reader will also think that something is wrong with the statement. Even though it makes sense on its own, the boss should not have written that statement in that specific context. Normally, a letter of recommendation should discuss the candidate's qualifications for the position, personal qualities that would make her a good employee, and so forth. Because good hair does not qualify Susan for a job (unless it's some kind of hair modeling job), the statement lacks relevance, and therefore constitutes a faulty or flawed contribution in that situation.

Philosopher of language H. Paul Grice analyzed conversations and exchanges like the one just sketched here to develop a theory of dialogue (Grice, 1989). His principles or

rules of conversation are known as *Gricean maxims*. These principles are not hard-and-fast, absolute prescriptions for verbal behavior (unlike some principles of grammar, which are rarely or never violated). Instead, they are a set of guidelines that speakers normally adhere to, but sometimes ignore. However, when speakers ignore one of the Gricean maxims, there is usually a reason why they do so. In Susan's letter of recommendation, the fact that her boss was only willing to comment on her hair probably tells us that Susan wasn't a very good employee. That is, the boss violated the principle that says "be relevant." In this case, what's relevant to the exchange is how good a worker Susan is. Because the boss does not comment on Susan's work habits, the obvious inference is that there isn't much good to say about them.[1]

Grice proposed that conversational principles should be organized in a hierarchy. The *cooperative principle* stands at the top of the hierarchy, and other principles are organized below that. According to Grice, the overarching goal of conversation is the exchange of information, and to do that speakers design their utterances to help the listener acquire information. That is, speakers cooperate with listeners whose goal is to learn something new. In Grice's terms (Grice, 1989, p. 26), the cooperative principle directs speakers to, "Make your conversational contribution such as is required at the stage at which is occurs, by the accepted purpose or direction of the talk exchange in which you are engaged." Speakers most often cooperate, although sometimes they lie or purposely seek to confuse or baffle their listeners.

To cooperate with listeners, speakers need to take into account listeners' presuppositions, beliefs, and knowledge. For example, if both partners in a conversation know that a person called Bob is very tired, it makes sense to say, *Bob succeeded in keeping his eyes open*. But if a listener thinks or knows that Bob is alert and chipper, then that same statement makes little or no sense. The speaker should choose the expression *Bob succeeded in keeping his eyes open* only if the speaker believes the listener thinks Bob is tired. To be cooperative therefore the speaker must have and use knowledge about the listener's mental state.

When speakers cooperate with listeners, they follow a number of additional principles that determine whether their statements are well-formed and contextually appropriate. The main principles are those of *quantity, quality, relation*, and *manner*.

Grice's *principle of quantity* says: "1. Make your contribution as informative as is required (for the current purposes of the exchange). 2. Do not make your contribution more informative than is required" (Grice, 1989, p. 26). Saying too little is costly, although it can be amusing, as when Graham Chapman addresses the question of how to rid the world of all known diseases (Python, 1990, p. 63):

> *Well, first become a doctor and discover a marvelous cure for something, and then, when the medical profession really starts to take notice of you, you can jolly well tell them what to do and make sure they get everything right so there'll never be any diseases ever again.*

Saying too much can also be costly. If someone asks you *How do I get downtown from here?* you do *not* say *Walk to your car, place your hand on the door handle. Raise the door handle until the door opens. Move your body so that you are sitting behind the steering wheel* ... This violation of quantity is costly because it delays communication of information that meets the listener's needs and clutters the listener's representation of the discourse with irrelevant detail.

Grice's *principle of quality* says: "Try to make your contribution one that is true ... 1. Do not say what you believe to be false. 2. Do not say that for which you lack adequate evidence" (Grice, 1989, p. 27). So, speakers should not lie or just make stuff up.

According to the *principle of relation*, a speaker's contributions should be relevant to the current topic of conversation. Listeners who assume that speakers are following the principle of relation can successfully interpret potentially ambiguous statements. Grice (1989, p. 89) provides the word *grass* as an example. If the topic of conversation is about landscaping, and the speaker says *This is really good grass*, the principle of relation dictates that the intended meaning has to do with the stuff that grows on the front lawns of houses in American suburbia. However, in the context of a late-night debate about national drug control policy, the principle of relation should cause listeners to favor the "marijuana" meaning of *grass*. It would be weird if your lawn guy started talking about marijuana. It would be equally weird if lawn care sprang up during a discussion of drug policy.

The Gricean *principle of manner* (Grice, 1989, p. 27) says: "Be perspicuous" (that is, clear in expression or statement[2]). To be perspicuous, a speaker should: "1. Avoid obscurity of expression. 2. Avoid ambiguity. 3. Be brief (avoid unnecessary prolixity). 4. Be orderly." Style manuals, such as Strunk and White, provide further detailed advice about how to achieve the Gricean ideal of manner.

Taken as a whole, Grice's maxims provide us with a recipe for accomplishing two separate but vital communicative functions. On the one hand, they illustrate factors that speakers should take into account to craft effective statements. On the other hand, they provide listeners with clues that enable them to draw inferences that go beyond the literal content of a speaker's statements. However, Grice seemingly never intended his principles to serve as a theory of how speech is actually planned under the time pressures that limit speakers' performance in real conversation, and, in fact, some aspects of real conversation appear inconsistent with Grice's principles.[3] Subsequent theorizing in psycholinguistics has attempted to deal with some of these shortcomings. While Gricean principles are still recognized as a description of ideal speech planning and inference in dialogue, more recent accounts try to accommodate a wider range of phenomena that occur in real dialogues.

Dialogue Is Interactive

The opportunity for interaction is one of the main factors that distinguishes dialogue from monologue.[4] Classical approaches to dialogue (e.g. Cherry, 1956), however, view each participant in a dialogue as operating independently, somewhat like chess players. In chess, you think and then you make a move. Then I think for a bit and make a move. Then you think some more and make another move. Each move depends on the move that precedes it, but the decisions that get made are made by each individual thinking and acting alone. A similar process in dialogue would start with me saying something and you listening. When I'm done, you say something and I listen. We go back and forth, each taking turns to speak. As Martin Pickering and Simon Garrod note, the classic view of dialogue "simply involves chunks of monologue stuck together" (Pickering and Garrod, 2004, p. 170).

In real dialogue, moves are not strictly sequential—the beginning of one person's speech often overlaps with the end of someone else's turn—and significant portions of the dialogue are created by partners interacting, cooperating, and collaborating on the content of the dialogue (Clark, 1996; Garrod and Pickering, 2004; Stivers et al., 2009). Further, rather than consisting of a series of one-off statements, dialogue is built up from multi-turn exchanges, as in the following example (Bangerter and Clark, 2003, p. 212; see also Clark and Schaefer, 1989). In this case, a director is attempting to tell a listener

how to put together a set of Lego blocks to make a specific figure (* marks indicate where the two participants were speaking simultaneously).

A1: Okay. Um let's see. So we need a yellow two by two. Okay and that's going to fit on the right side of the blue block.

B1: M-hm.

A2: So that half of it oh yeah on one row of the right side of the blue block.

B2: Okay *so half of it's pointing to the right.

A3: *So half of it is pointing off to the right. Yeah.

B3: Got it.

In this case, speaker A's chief goal is to get listener B to put together two Lego blocks in a particular way.

According to the classical theory of dialogue, speaker A would think about how to describe the arrangement of Lego blocks and produce a statement that communicates that arrangement. The listener would then decipher the speaker's statement in order to recover the information it conveyed. However, the transcript shows something more complex than that taking place. Speaker A breaks the superordinate goal into sub-goals and begins by conveying one of them (statement A1). Listener B signals acceptance of the first part of the message (statement B1). In statement A2, the speaker begins to convey the rest of the instruction, catches themself halfway through, and reformulates the message. At that point, listener B helps out by elaborating on the information given in statement A2, which speaker A simultaneously verifies by saying almost exactly the same thing in statement A3. The ability of listener B to help out speaker A indicates that listener B is actively anticipating where the dialogue is likely to go, suggesting both incredibly rapid interpretation of speaker A's previous statement as well as a highly accurate understanding of the speaker's point of view. The listener verifies the content of the entire exchange by saying *Got it* (B3), providing the speaker with concrete evidence that their attempts to communicate have succeeded.

Examples like these illustrate that, rather than being a sort of "ballistic," "all-or-nothing" activity involving independent moves by independent players, dialogue involves a great deal of collaboration and joint, cooperative activity. Contrary to the maxims of manner and quality, speakers produce tentative descriptions before they have fully worked out all the details of their messages or determined the most effective way to express their ideas. Listeners are not passive. They actively anticipate how the conversation will evolve, and they provide speakers with explicit evidence of their understanding via both *backchannel* responses (head nods, *mhmmmm* noises) and main channel responses (*What? Speak up, sonny! Got it*).

Common Ground

Dialogues from the Lego task, and others like it, show that speakers do not always craft perfectly Gricean utterances. Sometimes, speakers lack the information they need to be perfectly informative, clear, concise, and helpful. Sometimes, they have the information but do not have enough time to fully evaluate it before they begin to speak. Despite violating Gricean principles of dialogue, speakers routinely manage to communicate effectively with listeners in face-to-face exchanges. This high level of communicative success is enabled largely by opportunities for cooperation and collaboration that interactive dialogue provides.

Herb Clark and his colleagues have spent considerable effort building a theory of dialogue that attempts to capture the interactive nature of the communicative process (Clark, 1996; Clark and Schaefer, 1987; Schober and Clark, 1989; Wilkes-Gibbs and Clark, 1992). According to Clark and colleagues, the main goal in conversation is for partners to establish and expand *common ground.*

"Common ground" sounds an awful lot like "common knowledge," or "shared knowledge," but it is really something different. "Common knowledge" or "shared knowledge" implies that two people know the same thing. But two people can know the same thing without having any awareness that the knowledge is shared. Common ground is a kind of shared knowledge, but common ground requires something extra: both of the people in an exchange have to explicitly recognize that the knowledge is shared. So, "common *knowledge*" is "shared knowledge," but common *ground* is "mutually recognized common knowledge." Because of its emphasis on common ground, Clark's theory is sometimes referred to as *common-ground theory*, or the *common-ground theory* of dialogue.

According to Clark's common-ground theory, successful communication takes place when two people expand the amount of common ground that they share. This approach helps to explain why contributions to dialogue normally consist of more than just single statements (as in the Lego example earlier). Because the goal of conversation is to expand common ground, and because the contents of common ground must be acknowledged by each conversational partner, speakers require some kind of evidence that each of their statements have succeeded in adding to common ground.[5] Rather than just producing a sequence of statements, speakers in dialogues use information that is already in common ground to formulate statements, and they collaborate with their partners until both of them believe that communication has succeeded (Clark and Wilkes-Gibbs, 1986; Wilkes-Gibbs and Clark, 1992). Evidence of understanding can take a variety of forms, from implicit acceptance through backchannel responses (*mhmmmm*) to explicit statements of acceptance (*Got it*) (Bard et al., 2007).

Common-ground theory helps explain how partners in dialogue choose particular expressions to refer to specific concepts, and how referring expressions evolve over the course of multiple exchanges between conversational partners. Early in an exchange, conversational partners will often negotiate about how to refer to something, and they will each make several individual contributions to the conversation as they craft and agree upon a particular referring expression.

In a task where one person tries to describe an abstract picture to another, you might see an exchange like this the first time the two try to accomplish the task (from Wilkes-Gibbs and Clark, 1992, p. 184):

A: Okay the next one is … resembles someone that looks like they're trying to climb stairs. There's two feet, one is way above the other, and …

B: And there's a, there's a, a diamond on the right side, on a slant?

A: Yeah.

B: Got it.

A: Like, kind of looks like it's off the back.

B: Right, I got it.

But if the two are asked to do the same task, with the same pictures, a few more times, they will eventually refer to the same picture like this:

A: Uh, the next one is the person climbing the stairs.

B: OK.

And then like this:

A: Stair climber.

B: OK.

In the original exchange, the partners had to verify that they were looking at the same picture. In the second exchange, the speaker used part of the description from the first exchange (*someone that looks like they're trying to climb stairs*) to identify the picture, but can shorten the description because a reference to someone climbing the stairs has already been established in common ground (so the description becomes *person climbing the stairs*). In the final exchange, the description becomes shorter still.

Participants in dialogue routinely craft referring statements by collaborating and verifying mutual understanding, and once those referring statements have been established, both participants in the dialogue can use them without further elaboration (see also Barr and Keysar, 2002; Carletta et al., 1998; Fay et al., 2000). Further, if speakers violate the "referential pacts" that they negotiated previously with a specific listener (e.g. if speaker A in the previous exchange had suddenly started calling the *stair climber* the *diamond lady*), listeners experience some difficulty figuring out what the new term refers to (Metzing and Brennan, 2003). Thus, because prior exchanges establish information in common ground, a fuller description of the physical appearance of the picture is unnecessary, and communication becomes more efficient.

Because different pairs of conversational partners have different common ground, speakers need to pay attention to who, exactly, they are speaking to so that they can craft optimally effective messages. If I assume that a particular piece of knowledge is in common ground with a listener, but that knowledge is not actually in common ground, my attempt to communicate is likely to fail. However, once you and I have established common ground, we can exploit that resource to communicate more effectively with one another *and* to keep information out of the hands of people who overhear our conversation (Clark and Schaefer, 1987; see also Schober and Clark, 1989).

In experiments involving pairs of friends conversing with one another, the conversational partners took advantage of personal experiences to craft referring expressions that unfamiliar third persons had difficulty interpreting. In these experiments, pairs of friends were asked to describe common landmarks from their university campus (such as a central fountain). Because pairs of friends had private experiences related to those landmarks, and those private experiences were part of the common ground between them, they could use those experiences (e.g. *This is where I put your teddy bear*) rather than the publicly available labels for the landmarks. While there was some "leakage" of private information, pairs of friends enjoyed great success concealing the true identity of the landmarks when they were instructed to do so (compared to when they were simply asked to describe the landmarks, without any reference to possible third parties). This success in communicating to each other and concealing information from third parties happened because two friends could take advantage of common ground that they shared. The third party failed to understand much of the conversation because the third party lacked the crucial knowledge supplied by common ground.

Common ground theory further offers a fresh perspective on a couple of old questions in dialogue research: How do people decide when it's time to take a turn in conversation? And why do speakers go *uhhhhhhhhhh*? According to one influential analysis of turn-taking behavior in conversation (Sacks et al., 1974), who gets to talk when is determined by a set of "soft" constraints (as opposed to hard-and-fast rules), which, in combination, influence when people start and stop speaking during conversation. For example, in a group discussion, the person who is speaking now can influence who will

speak next by looking at one specific person. The person that the speaker looks at is most likely to take the next turn (although someone else certainly can jump in). Likewise, speakers are likely to look off into space if they plan to continue to speak for a while, but they will look at someone if they think their turn is coming to an end. Another rule specifies that overlap between speakers must be minimized—in most cultures, conversations are carried out by one person speaking while everyone else listens (Stivers et al., 2009). But if the speaker cannot immediately think of what to say next, and therefore pauses in the middle of their turn, someone else could jump in and cut the speaker off before they have said what they wanted to say.

One theory of why speakers go *ummmmmmmm* is that they are trying to eliminate pauses in their speech, and thereby hold the floor until they have said what they want to say. Speakers can perform a similar trick by lengthening the pronunciation of function words, for example, by pronouncing the word *the* as *theeeeeee* or *a* as *uhhhhhhhhhh* when they want to hold the floor as they are trying to think of what to say next. So, in classical theories of dialogue, sounds like *uhhhhhhhh* and *theeeee* are called *filled pauses*. Alternatively, *theeeee* and *uhhhhhhhh* have been viewed as simple production errors. (Evidence against the "*uhhhhhhhh*-as-error" hypothesis comes from the fact that drunk people say *uhhhhhhhh* less than sober people. That's because they care less about making sense, much less holding the floor (Christenfeld and Creager, 1996).)

According to Clark and colleagues (Brennan and Williams, 1995; Clark and Fox Tree, 2002; Fox Tree, 2001; Fox Tree and Clark, 1997), however, *theeeee* and *uhhhhhhhh* are conventional words in and of themselves, just like *cat* and *house*. Just like other words, speakers have control over when they produce *theeeee* and *uhhhhhhhh*, and these words serve to place particular concepts into common ground. Specifically, *theeeee* and *uhhhhhhhh* place into common ground knowledge that the speaker is experiencing an "upcoming delay (in speech planning) worthy of comment" (Clark and Fox Tree, 2002, p. 73). To provide evidence for that claim, Clark and colleagues can point to samples of naturally occurring speech containing instances of people saying *theeeee* and *uhhhhhhhh*. In these speech samples, speakers actually produce different versions of *uhhhhhhhh*, a short version—more like *um*—and a long version—*uhhhhhhhh*. In the speech samples, *um* is usually followed by a short pause, while *uhhhhhhhh* is followed by a longer pause. Because *um* appears before short pauses and *uhhhhhhhh* before long pauses, it must be the case that speakers have the ability to anticipate how long the pause is going to be (otherwise, the distribution of *um* and *uhhhhhhhh* would be random). Further, if *theeeee* and *uhhhhhhhh* are real words, then listeners should go beyond the speaker's explicit statement and draw inferences when *theeeee* and *uhhhhhhhh* appear in a statement. One inference that listeners draw from the long version, *uhhhhhhhh*, is that speakers are experiencing trouble recovering a word from memory. Under those circumstances, listeners are more likely to jump in and help complete the speaker's statement, and they are more likely to pay attention to unfamiliar objects in a visual display (Arnold, Tanenhaus, et al., 2004; Arnold et al., 2007).

Audience Design

Herb Clark's *common ground* approach proposes that speakers in dialogues commit significant mental resources to modeling their listeners' knowledge states, specifically by keeping track of the information that the speaker has in common ground with specific listeners. Speakers then use this knowledge to plan statements so that their

listeners have the best possible chance of understanding them, for example, by using referring expressions that have been negotiated previously. Clark's theory of dialogue therefore incorporates the idea that speakers routinely engage in *audience design*: they take special pains to adapt what they say to meet the listener's specific needs. This does not mean that speakers consciously evaluate every possible way to craft their message. Speakers certainly *can* engage in painstaking, conscious deliberations, because deciding what to say and how to say it are under some degree of conscious control. But Clark's theory says that knowledge of the listener's needs can be rapidly and accurately deduced from a representation of common ground, and this can be done outside of conscious awareness and control. In fact because speakers are under considerable time pressure during normal conversations, there may not be enough time for painstaking, exhaustive analysis, and so production processes may rely largely upon rapid, relatively automatic processes.

In general, speakers *do* appear to make some adjustments depending on who they are speaking to: Bilingual speakers can change languages. We talk differently to babies than to adults. If someone appears to be from out of town, we give more detailed directions than if they appear to be from the same city as us. In a noisy environment, we talk louder so the listener can hear better. These adjustments reflect a kind of *macro-audience design* where speakers make a decision at one point and apply that decision over an extended period of time. The common-ground approach and others like it propose that speakers engage in a more *micro* style of audience design, where they are continuously making small adjustments based on moment-to-moment changes in the listener's knowledge, attentional state, and needs (Hanna et al., 2003; Hawkins et al., 2021; Levelt, 1983). While there is general agreement that speakers can and do make macro-level adjustments when they speak, there is considerably less agreement about the extent to which speakers make micro-level adjustments.

Other theories of how speakers plan utterances view speakers as being more *egocentric* (Ferreira and Dell, 2000; Horton and Gerrig, 2005a, 2005b; Horton and Spieler, 2007; Keysar and Henly, 2002). That is, rather than dedicating mental resources to modeling listener's knowledge and tailoring utterances to meet the listener's specific needs, speakers' utterances are driven more by *availability*. The principle of availability says that speakers will produce whatever is easiest for them to produce, even when what is easy for speakers to say may be more difficult for listeners to understand. Information that is highly active, salient, or *available* to speakers will be easiest to produce, and so that information will show up in utterances. If that information is unhelpful to a listener, the speaker and listener will engage in a process of negotiation and repair to get communication back on track (as in the *common ground* approach).

Accounts such as Keysar and colleagues' *monitoring and adjustment* hypothesis fall within the *egocentric production* class of theories because they propose that speakers ignore listeners' knowledge and needs during initial utterance planning. However, speakers do monitor their own utterances, as well as feedback from listeners, and can adjust their messages based on the outcome of the monitoring process. But any such statements will appear relatively late and speakers are likely to produce utterances that are suboptimal from the listener's perspective some of the time—or perhaps even quite often (Lau et al., 2022).

It's important to study whether speakers in dialogue behave egocentrically or whether they design their utterances for their listeners, because determining the extent to which speakers behave egocentrically can help us figure out which theory of dialogue is the right one. As a result, researchers have spent considerable effort investigating whether speakers are egocentric.

Dialogue research has enjoyed an upsurge of interest, in part because advances in head-mounted eye-tracking technology have opened up new avenues of research. These experiments sometimes involve a real participant and a *confederate*, or *stooge*—someone who is working directly for the experimenter and is following a tightly controlled script. Having a stooge in the experiment helps experimenters get control over the *exuberant-responsing* problem in dialogue. That is, people in dialogues can say whatever they want, whenever they want to, and this makes it hard to get control over variables that researchers might be interested in testing. Having a stooge means that researchers can present specific utterances at specific times, and they can then see how real participants react. Other experiments involve pairs of real participants, none of whom works for the experimenter. These latter experiments have the advantage of providing genuine feedback and interaction, at the cost of losing some of the experimental control that stooges provide.

If speakers engage in (micro-level) audience design, then they should avoid producing utterances that may cause difficulty for their listeners when an alternative is available to them. Syntactically ambiguous utterances are often difficult for people to understand. At least, they are more difficult to understand than equivalent *un*ambiguous utterances. If speakers take their listeners' needs into account when they are planning what to say, they should avoid syntactically ambiguous utterances. That being the case, speakers should be more likely to say something like (1b) than (1a):

(1) a. I knew the coach and his sister would arrive late.
b. I knew <u>that</u> the coach and his sister would arrive late.

Inserting the optional word *that* in (1b) renders the sentence unambiguous, and sentences like (1b) really are a little easier to understand than ambiguous sentences like (1a).

Vic Ferreira and Gary Dell tested whether speakers avoid ambiguous expressions by having participants read sentences such as (1a) and (1b) (Ferreira and Dell, 2000).[6] Later on, they repeated them back from memory (see Arnold, Wasow et al., 2004; Ferreira et al., 2005; Haywood et al., 2005; Keysar and Henly, 2002; Kraljic and Brennan, 2005; but see also Lockridge and Brennan, 2002). The question was whether speakers would add the optional word *that*, which would indicate that they were trying to make their sentences less ambiguous and therefore easier to understand.

In this experiment, speakers did sometimes include the optional word *that*, but they did so to the same degree whether the word *that* was necessary to render the sentence unambiguous or not (sometimes, other cues meant that the sentence was already unambiguous, even without the word that). Speakers did not appear to avoid ambiguity, and many of their utterances were, in fact, ambiguous. When speakers did use optional words that *could* make the sentence unambiguous, they did so just as much when the sentence was already unambiguous as they did when adding the word *that* could actually help out a listener. What really determined whether speakers used the word *that* was how hard it was for them to access the next word (e.g. *coach*). If a word had been recently encountered or was otherwise made highly salient, speakers left out the word *that* when they produced the sentence. Ferreira and Dell concluded that speakers use optional words such as *that*, not because they make life easier for listeners, but because they make life easier for speakers. When speakers are searching memory for the next word to say, and when that search is going a little bit slow, speakers can insert the highly frequent, closed-class, function word *that* to buy themselves just a little bit of extra time. (As Vic once said, saying *that* is a fancy way of going *uhhhhhhhh* …)

Dialogue offers speakers other ways to engage in audience design besides using optional function words. When speakers are describing events, they make choices about how much detail to include (see Gricean maxim of *quantity*, earlier). If speakers engage

in audience design, they should include information that will be difficult or impossible for their listeners to infer, but they should omit or neglect to mention information that is obvious or that would be fairly easy for listeners to infer. Is this how speakers really behave?

To find out, Paula Brown and Gary Dell asked speakers to repeat short stories such as (2). The stories either mentioned a typical instrument (knife) or an unusual instrument (ice pick) (Brown and Dell, 1987, p. 444):

(2) The robber hid behind the door and when the man entered the kitchen he stabbed him in the back. He wiped the blood off the {knife/ice pick} and rummaged through the drawers. Later police investigators found his fingerprints all over the {knife/ice pick} and had no trouble catching him.

The participants' task was to read and retell the story from memory. If speakers engage in audience design, they should mention the ice pick more often than they mention the knife, because listeners are unlikely to infer that a strange instrument was used to stab the victim in the story. However, speakers might mention the ice pick more often simply because it is weird to stab someone with an ice pick, even more weird than stabbing them with a knife, and people in general tend to pay more attention to weird things than normal things.[7] *Ice pick* should be more salient and more memorable to the speakers themselves. So, rather than including *ice pick* when they retell the story because it helps listeners, speakers might mention *ice pick* just because it sticks in their mind. In fact, speakers do mention weird instruments like ice picks more than they mention normal instruments like knives, but this does not, by itself, tell us what causes speakers to behave this way.

To find out whether audience design or salience better explains why speakers mention weird instruments, speakers and (confederate) listeners viewed pictures such as the one in Figure 8.1 as the speakers repeated the stories. In the control condition, no instrument appeared in the picture. Having the picture present for both the speaker and the listener establishes a kind of common ground. When the left-hand picture is used, both the speaker and the listener know that an ice pick will be relevant to the story even before the speaker begins to tell the story, so the ice pick is part of the common ground. When

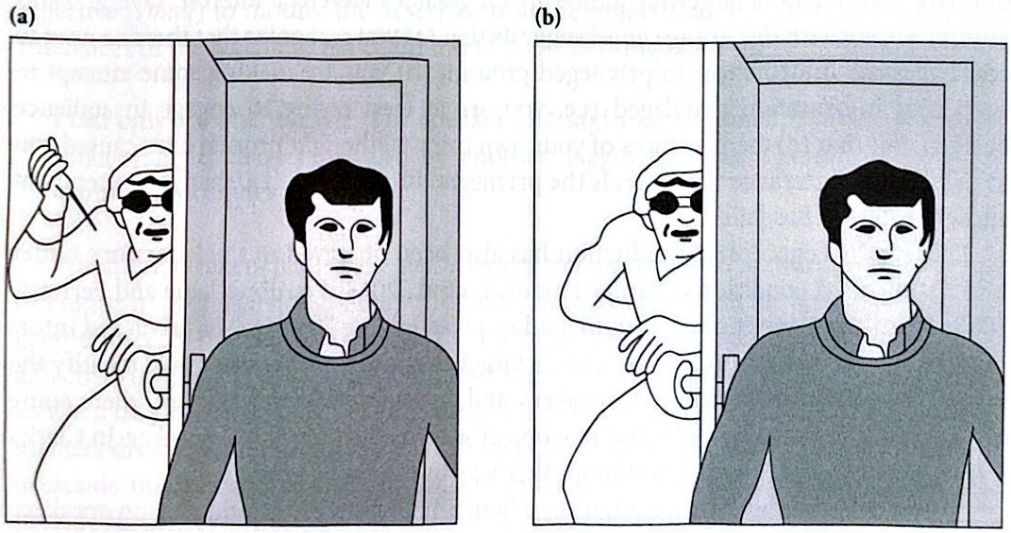

Figure 8.1 Picture depicting "weird" ice pick instrument (left) or no instrument (right; the control condition). *Source:* Brown and Dell (1987), with permission of Elsevier

the right-hand picture is used, only the speaker knows that an ice pick will be relevant to the story. So the ice pick is part of the speaker's privileged ground. If common ground is used by speakers to design their utterances, they should mention the ice pick more often when the picture fails to provide the necessary information. In fact, speakers were just as likely to mention the ice pick when the listener was looking at the left-hand (ice pick) picture as when the listener was looking at the right-hand picture. Rather than mentioning the ice pick because listeners needed that verbal information to understand the story, speakers mentioned the ice pick just because it is a strange and attention-grabbing element of the story. In other words, speakers in this experiment were behaving egocentrically.

Egocentric production

Three may keep a secret if two of them are dead.
BENJAMIN FRANKLIN

In the preceding example, mentioning the ice pick when the listener already knows about the ice pick (by virtue of having a picture) does not help listeners understand the story, but it certainly does not hurt anything. Speakers are mildly violating Grice's maxim of quantity, but the cost that this violation creates for speakers and listeners is really very small. There are other circumstances, however, where failing to engage in audience design can impose costs on the speaker. When failing to take the listener's knowledge into account creates problems for the speaker, we would expect speakers to make special efforts to engage in audience design.

Consider a situation where you have a big secret. Let's say you're a waiter and someone ordered orange juice. Let's say you found something in the orange juice, but you decide to serve it anyway. If you want a nice tip, what you do *not* want to do is alert the customer to any possible problems with the orange juice. So you should say something like *Here is your orange juice*, or just *Here you go*. But if you are thinking about what you just found in the orange juice, if that really captured your attention, you might actually say, *There's nothing wrong with this orange juice*. If that happened, then the salience of your private thoughts would have trumped your desire to form an utterance that would improve your chances of getting a nice tip. Of greater theoretical interest, saying *There's nothing wrong with this orange juice*[8] suggests that (a) you recognize that there's a need to conceal some information in privileged ground, (b) you are making some attempt to keep that information privileged (i.e. you are at least trying to engage in audience design), but that (c) the pressures of your own internal thought process have caused you to produce an utterance that reveals the privileged information (i.e. that your attempt at audience design has failed).

This kind of egocentric production has also been observed in the laboratory under more controlled conditions (Vanlangendonck et al., 2018; Wardlow-Lane and Ferreira, 2008; Wardlow-Lane et al., 2006). In studies investigating "leakage" of privileged information, speakers were asked to describe a target shape so that listeners could identify the target. As shown in Figure 8.2, speakers and listeners viewed a display where some objects could be seen by both, but one object was visible only to the speaker. In Clark's terms, the object visible only to the speaker was in *privileged ground*.

To test whether audience design or salience had a stronger influence on speakers' behavior, the experimenters placed audience design and salience into conflict. The experimental task required speakers to describe one of the mutually visible objects (the *target*), so that the listeners could identify it. On some of the trials, the object in privileged

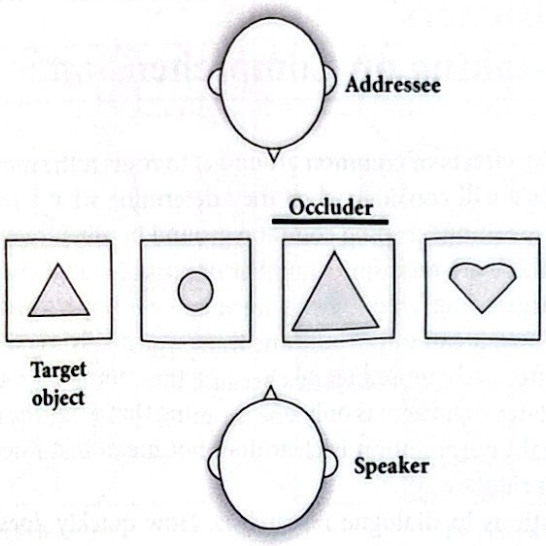

Figure 8.2 Experimental set-up. The "occluder" prevents the listener ("addressee") from seeing one of the objects. The speaker's job is to get the listener to identify the target object, without being able to guess what is behind the occluder. *Source:* Wardlow-Lane et al. (2006), SAGE Publications

ground was the same shape as the target object. The tricky part was that the speakers were supposed to keep the identity of the privileged object a secret. If the speakers engaged in audience design, that is, if they considered only the objects that were in common ground (everything except the occluded object), then they would realize that the listeners could see only three shapes, and the three shapes were unique (e.g. a triangle (target), a circle, and a heart). Thus, a speaker could get the listener to identify the target simply by saying, *Pick the triangle*. Note, however, that from the speaker's own perspective, *Pick the triangle* is ambiguous, because both the target object and the privileged object meet that description (they are both triangles). So, if the speaker designs an utterance taking into account only their own perspective, and not the listener's, the speaker might very well say, *Pick the small triangle*. However, as soon as the speaker uses an adjective (*small*) to modify the description of the object that should set off a chain of inferences in the listener. That chain might go something like this:

> I can only see one triangle. The speaker just said *Pick the small triangle*. If there's a small triangle, there must also be another, bigger triangle. Therefore, the object behind the occluder is a triangle.

When the speakers fail to take the listeners' perspective into account, when they fail to engage in audience design, they will make mistakes (because the speakers' task is to keep the secret information secret). In this experiment, speakers frequently included adjectives (e.g. *small*) when describing the target objects, and they were *more* likely to do so when they were instructed to keep the occluded object secret. Thus, when the experimenters drew attention to the occluded object, the salience of the object to the speaker overcame the speaker's abilities to engage in audience design. Experiments like these provide further evidence that the state of speakers' internal mental representations, in particular the degree to which specific concepts are attentionally focused or accessible, substantially affects production processes and may short-circuit speakers' attempts to design utterances for particular listeners.

Effects of Listeners' Perspective-taking on Comprehension

Recall that one of the effects of common ground is to restrict the number of things that a speaker and listener will consider when they determine what a referring expression means (technically speaking, keeping common ground in mind "restricts the domain of reference"). While there are an infinite number of possible kings (King Henry V, Good King Wenceslas, Burger King™, etc.), if you are at an Elvis impersonator convention and someone says *The King*, it can only mean one thing. At the Elvis convention, the meaning of *The King* is effectively unambiguous because the common ground available to all of the attendees dictates that there is only one meaning that would be relevant. However, just because the final interpretation is clear does not mean that listeners automatically access that meaning right away.

One of the questions in dialogue research is: How quickly does common ground restrict the domain of reference? Does common ground automatically and immediately render referring expressions like *The King* unambiguous? Or do participants in conversations need to go through a process of disambiguating referring expressions, the same way they go through a process of disambiguating words like *bank*, and *mint* that have more than one meaning? Another way to frame that question is to look at it through the prism of common ground versus egocentric styles of processing. If listeners engage in egocentric processing when they interpret speakers' statements, then they may come up with meanings that the speaker did not intend. By contrast, if speakers make use of common ground when they plan their statements, and listeners also keep track of what concepts are included in common ground, then listeners should immediately interpret speakers' statements as the speakers intended.

Some experiments provide strong evidence in favor of the egocentric listener hypothesis (Keysar et al., 2000, 1998; Wu and Keysar, 2007; see also Corps et al., 2022; Kronmüller and Guerra, 2020). In these experiments, a speaker described objects to a listener and asked the listener to move the objects around a vertical grid (see Figure 8.3 for an example). The speaker sat on one side of the grid, while the listener sat on the other. To manipulate privileged and common ground, some of the objects were visible to both participants, but some were visible only to the listener. In Figure 8.3, the speaker could see a bottle, a truck, and two candles. Importantly, the listener could see *three* candles, the two that were visible to the speaker, plus another one. The smallest candle therefore was in the listener's privileged ground, and was definitely *not* a part of the common ground. (In this case, common ground is defined as the set of objects that are visible to both the speaker and the listener.) The question that Boaz Keysar and his colleagues asked was: "If the speaker says *Pick up the small candle*, what will the listener do?" The listener's behavior depends on whether the listener consults common ground before she decides what the term *the small candle* refers to. If the listener consults common ground (if common ground restricts the domain of reference), she will find only two candles there (the one on the top row, and the bigger of the two candles on the bottom row). If those two candles make up the *contrast set* (the total set of objects that can be referred to with the term *candle*), then the term *the small candle* should be interpreted as meaning the left-most candle on the bottom row in Figure 8.3. But what if the listener ignores common ground? If the listener ignores common ground, then there are going to be three candles in the contrast set. That is, the set of potential referents for *candle* will include even the very tiny candle that the speaker cannot see. As a result, an egocentric listener—one who ignores common ground—will interpret *the small candle* as meaning the smallest one the listener can see (i.e. the right-most candle in Figure 8.3).

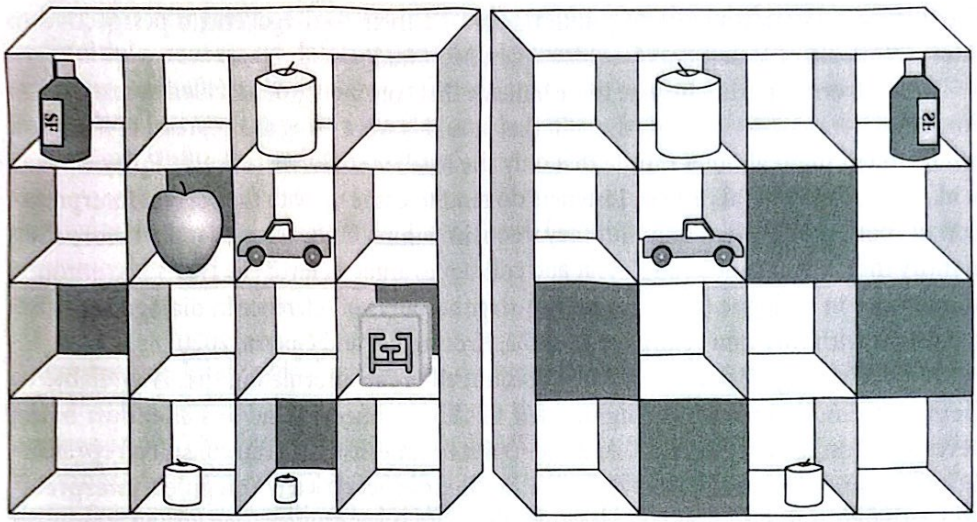

Figure 8.3 Examples of grids. The listener's (addressee's) view is shown on the left. The director's view is on the right. Note that some of the objects were visible to both participants (e.g. the truck), but some were visible only to the listener (e.g. the apple, the block, and the smallest of the three candles). *Source:* Keysar et al. (2000), SAGE Publications

In this experiment (Keysar et al., 2000), listeners wore a head-mounted eye-tracker so that the experimenters would know where listeners were looking as they listened to instructions such as *Pick up the small candle*, spoken by the confederate. Because there is a reasonably tight linkage between what people are thinking about and where they are looking (Tanenhaus et al., 1995), researchers can deduce which meaning listeners assigned to *the small candle* by checking where they look when they hear the word *candle*.[9] If listeners look at or pick up the smallest candle in their display (the one that speakers cannot see), this would mean that they are ignoring common ground while they are figuring out what the expression *the small candle* refers to. In fact, listeners in this experiment and others like it, frequently looked at the very smallest candle, and sometimes they picked it up and moved it, before they looked at the correct candle: listeners behaved egocentrically.

In an even more extreme version of the "candle" experiment, let's call it the "tape in a bag" experiment, listeners interacted with a confederate speaker (Keysar et al., 2003). The task was very similar to the "candle" experiment in that listeners responded to instructions spoken by the confederate. A critical aspect of this study was that, on some trials, without the knowledge of the speaker, the listener hid an object inside a paper bag. For example, the listener hid a roll of Scotch Tape™ inside a paper lunch bag. Then, they placed the paper bag (with the tape hidden inside it), on a grid just like the grid in the "candle" experiment. The grid contained other objects, including a cassette tape. The vital question in this experiment was how listeners would react when the speaker said *Move the tape*. If listeners consider common ground when interpreting *tape*, then *tape* can only refer to the mutually visible cassette tape. But hiding the Scotch Tape in the bag makes that potential referent very highly salient to listeners. Egocentric pressures might therefore cause listeners to think of Scotch Tape when a speaker says *tape*, even though there is no way that the speaker can know what is inside the bag. In fact, listeners' overt behavior (the object that they reached for when they heard the word *tape*) as well as their eye movements (they tended to look at the paper bag when they heard *tape*) both indicated

that listeners ignored common ground and used their own egocentric perspective to interpret speakers' statements (see also Epley, Morewedge et al., 2004).

Results from experiments like these indicate that common ground failed to restrict the domain of reference—listeners thought that speakers were using the expression *the small candle* to tell them about a candle that only the listener could see (see also Epley, Keysar et al., 2004; Keysar et al., 1998). Listeners do tend to come up with the correct interpretation of *small candle* eventually, but they frequently think of the wrong candle before they identify the correct candle, so Keysar and colleagues argue in favor of a two-stage account of reference in dialogue (more recent *two-stage* accounts of reference in dialogue are also consistent with this view; Corps et al., 2022; Kronmüller and Guerra, 2020).

In the first stage, listeners take into account all of the information that is available to them, with more attention being devoted to salient concepts, and less attention being devoted to less salient concepts. At a later point in time, listeners can consult a representation of common ground to verify whether their egocentrically computed interpretation is likely to be correct. In cases like the candle experiment, consulting common ground will indicate to listeners that their initial interpretation is likely to be mistaken (because the speaker could not refer to an object that the speaker does not know about), and so the listener will come up with a revised interpretation that relies more on common ground than on egocentric sources of information. This aspect of Keysar and colleagues' account is known as the *perspective adjustment model*, which highlights the idea that listeners start out interpreting expressions based on their own egocentric perspective, and then make adjustments to that interpretation when they detect differences between their own perspective and the speaker's perspective.[10] The idea that adjustment away from your own perspective takes time receives support from experiments showing that listeners become more egocentric in their interpretations the less time they have to interpret referring expressions (i.e. when they have to come up with an interpretation before an artificial deadline; Epley, Keysar et al., 2004).

The preceding experiments suggest that listeners often interpret speakers' statements from an egocentric point of view, and take into account common ground with the speaker only after a delay. However, additional studies indicate that egocentric interpretation is not an inevitable and unavoidable aspect of dialogue. Cultural characteristics appear to affect the extent to which listeners engage in egocentric interpretation (Wu and Keysar, 2007). In experiments that compare American college students to exchange students from more collectivist cultures,[11] students from collectivist cultures behaved as though they consulted common ground very early during the process of interpreting speakers' statements. When they participated in the "tape in a bag" experiment, students from collectivist cultures rarely looked at the paper bag, and they never attempted to pick up or move the bag when the speaker said *move the tape*. Some studies of American adults and children also indicate that listeners do sometimes consult common ground fairly quickly to determine what speakers are referring to (Brown-Schmidt et al., 2008; Brown-Schmidt and Tanenhaus, 2008; Hanna and Tanenhaus, 2004; Hanna et al., 2003; Nadig and Sedivy, 2002). For example, in a task where a listener collaborated with a confederate speaker in a cooking task, listeners flexibly expanded or contracted the domain of reference depending on whether the speaker's hands were empty or full. When the speaker's hands were full, which meant that the speaker could not reach objects in front of them, listeners would consider objects in front of the speaker as potential referents for ambiguous expressions (e.g. *pick up the salt* when there were two salt shakers, one in front of the speaker and one in front of the listener). When the speaker's hands were empty, listeners would only consider objects that were in front of themselves. Mood also appears to affect the degree to which comprehenders engage in egocentric processing (Converse et al., 2008). People who are in a sad mental state appear to be less egocentric than people who are happy. Based on these kinds of results, some researchers

believe that interpreting references in dialogue results from the weighting and evaluation of multiple sources of information simultaneously. According to *constraint-based accounts*, egocentrically available information is one source of information that can influence a listener, common ground is another source, but neither source automatically has a bigger influence on interpretation.

Just what do you think you're doing, Dave? Dave, I really think I'm entitled to an answer to that question
HAL

Now that you know a bit about human dialogue, let's revisit for a moment the question of LaMDA's performance. Is LaMDA doing the same thing that people do? The short answer is no. The longer answer is that LaMDA can produce some very sophisticated responses that appear very similar to human responses, but does so via processes that differ from human cognition. LaMDA is sophisticated kin to Kurt Burgess' HAL and the Colorado group's LSA programs (see Chapter 3). It accesses huge databases of real human language by collecting samples from the internet. It performs computations over those samples to determine how likely different sequences of words are in different contexts. Some computer scientists argue that LaMDA's abilities are more closely related to what parrots do than what humans do. "Such systems (are) 'stochastic parrots'—they stitch together and parrot back language based on what they've seen before, without connection to underlying meaning" (Gebru and Mitchell, 2022). Remember the *grounding problem*? Looks like we have that with LaMDA as well.

Cognitive scientist Gary Marcus has always been generous in sharing his thoughts. Let's give him the last word on LaMDA:

> Neither Lamda nor any of its cousins … are remotely intelligent. All they do is match patterns, draw from massive statistical databases of human language. The patterns might be cool, but language these systems utter doesn't actually mean anything at all. And it sure as hell doesn't mean that these systems are sentient. (Marcus, 2022)

Summary and Conclusions

Dialogue is more than just chunks of monologue bolted together. Participants in dialogue simultaneously play the roles of speaker and listener—we monitor our own speech to make sure it comes out right and we plan what we are going to say next while we listen to other people speak. Although real dialogue often deviates from the ideals laid out by philosophers such as Paul Grice, and although real dialogue can be much harder to capture in the lab than monologue production or comprehension processes, dialogue follows lawful and consistent principles that we can discover via careful analysis. So, while speakers are not always optimally cooperative, perspicacious, and relevant, there are often good reasons why their behavior deviates from the ideal. Savvy listeners can capitalize on these deviations from the standard to read between the lines and accurately picture the speakers' true thoughts. More contemporary approaches to dialogue attempt to capture that interactive nature of the activity, and describe how interaction leads to expanded common ground and mutual knowledge. These interactive accounts capture features of dialogue that "serial monologue" accounts miss, such as why speakers are so heavily dependent on feedback, as well as helping explain how people manage turn-taking behavior. The common ground approach to dialogue naturally incorporates the

concept of audience design to explain how and why speakers make particular choices as they craft their contributions and how and why listeners develop specific interpretations of those contributions. While a substantial body of evidence suggests that, at the micro-level, speakers are often driven more by internal needs and pressures than by optimal attention to their listeners' needs, speakers do appear to routinely adjust at more macro-levels (and some recent experiments suggests that speakers can fine-tune their utterances at least some of the time). Likewise, considerable research efforts have been dedicated to listeners' processing of speakers' utterances in dialogue. Much of this research supports the idea that listeners are egocentric, at least during the early stages of comprehension, but some studies have shown that listeners adjust relatively quickly to specific knowledge about speakers. Further research will be required to identify factors that determine how and when listeners are able to overcome their own egocentric perspective to more quickly and accurately interpret a speaker's intended meaning.

TEST YOURSELF

1. Describe Paul Grice's conversational maxims and explain how they influence the way dialogues unfold. Why might Grice's ideas fail to explain some aspects of real conversations?
2. How do partners in a dialogue establish, maintain, and expand common ground?
3. How egocentric are speakers and listeners during dialogue. Describe experimental results that support or challenge egocentricity.
4. Why do speakers go *ummmm* and *ah* during conversation?
5. Explain the concept of "audience design," give an example, and describe experimental results that suggest speakers engage in audience design during conversation. Describe experimental results that suggest that they don't.

THINK ABOUT IT

1. Can you hold a dialogue with a partner who is *not* following Grice's maxim of cooperation? Why or why not? What would such a dialogue look like?

Notes

1 Of course, the boss could simply come out and say *Susan was a bad worker. Don't hire her*. Maybe the boss is trying to be polite ("If you can't say something good, don't say anything at all"). Maybe the boss is trying to avoid a lawsuit for defamation or slander.
2 Grice is a major violator of his own principles.
3 Research also calls into question the extent to which listeners expect speakers to be cooperative. Listeners appear to prefer consistent referring expressions, even in contexts where being consistent entails being less than optimally cooperative (Shintel and Keysar, 2007).

4 Grice never really considered how interaction and feedback might influence a speaker's behavior in dialogue. Grice argued that speakers craft utterances to produce a particular response in a target audience, but he does not talk about how speakers might react to specific feedback from that target audience. In his view, either the speaker would succeed in crafting an utterance or not. The listener played no active role in shaping a speaker's contribution.

5 Of course, different communicative situations involve different opportunities for interaction and grounding, and so speakers can adjust the extent to which they rely on feedback during utterance planning and execution. Written communication offers no opportunity for immediate feedback, a lecture offers some opportunity to sample the audience's response, while dialogue offers a multitude of immediate cues to communicative success. Speakers can continue to produce statements in the absence of feedback, but the likelihood of expanding common ground diminishes with diminishing feedback.

6 In the experiments of Haywood et al. (2005), naive participants were more likely to produce a disambiguating *that* when a confederate had just produced one and when a visual display provided a context in which an utterance lacking the word *that* would be ambiguous. Because no time-course information is available from that study, it is not clear whether visual context led to relatively immediate effects on production, or whether speakers engaged in some amount of self-monitoring, which could lead to increased detection of potential problems in cases where the visual context created an ambiguity of reference.

7 That's why you get wall-to-wall TV coverage of things like Falcon the Balloon Boy.

8 This really happened.

9 They can also check to see which objects listeners actually reach for and/or move. The results are the same whichever measure you prefer: Listeners frequently move the very smallest candle, the one that speakers cannot see.

10 In many cases, the speaker's and listener's perspectives will be tightly aligned, and so no adjustment will be necessary. In cases where the speaker's and listener's perspectives are different, feedback and collaboration can repair the misalignment.

11 *Collectivist* cultures place relatively greater emphasis on social relationships and group membership. *Individualist* cultures place relatively greater emphasis on independence and a sense of the self separate from others.

References

Arnold, J.E., Hudson Kam, C.L., and Tanenhaus, M.K. (2007). If you say Thee uh you are describing something hard: The on-line attribution of disfluency during reference comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 914–930.

Arnold, J.E., Tanenhaus, M.K., Altmann, R.J., et al. (2004). The old and thee, uh, new: Disfluency and reference resolution. *Psychological Science*, *15*, 578–582.

Arnold, J.E., Wasow, T., Asudeh, A., et al. (2004). Avoiding attachment ambiguities: The role of constituent ordering. *Journal of Memory and Language*, *51*, 55–70.

Bangerter, A., and Clark, H.H. (2003). Navigating joint projects with dialogue. *Cognitive Science*, *27*, 195–225.

Bard, E.G., Anderson, A.H., Chen, Y., et al. (2007). Let's you do that: Sharing the cognitive burdens of dialogue. *Journal of Memory and Language*, *57*, 616–641.

Barr, D.J., and Keysar, B. (2002). Anchoring comprehension in linguistic precedents. *Journal of Memory and Language*, *46*, 391–418.

Brennan, S.E., and Williams, M. (1995). The feeling of another's knowing: Prosody and filled pauses as cues to listeners about the metacognitive states of speakers. *Journal of Memory and Language*, *34*, 383–398.

Brown, P.M., and Dell, G.S. (1987). Adapting production to comprehension: The explicit mention of instruments. *Cognitive Psychology*, *19*, 441–472.

Brown-Schmidt, S., Gunlogson, C., and Tanenhaus, M.K. (2008). Addressees distinguish shared from private information when interpreting questions during interactive conversation. *Cognition*, *107*, 1122–1134.

Brown-Schmidt, S., and Tanenhaus, M.K. (2008). Real-time investigation of referential domains in unscripted conversation: A targeted language game approach. *Cognitive Science*, *32*, 643–684.

Carletta, J., Garrod, S., and Fraser-Krauss, H. (1998). Placement of authority and communication patterns in workplace groups. *Small Group Research*, *29*, 531–559.

Cherry, E.C. (1956). *On human communication*. MIT Press.

Christenfeld, N., and Creager, B. (1996). Anxiety, alcohol, aphasia, and ums. *Journal of Personality and Social Psychology, 70*, 451–460.

Clark, H.H. (1996). *Using Language*. Cambridge University Press.

Clark, H.H., and Fox Tree, J.E. (2002). Using uh and um in spontaneous speaking. *Cognition, 84*, 73–111.

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

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

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

Converse, B.A., Lin, S., Keysar, B., et al. (2008). In the mood to get over yourself: Mood affects theory of mind use. *Emotion, 8*, 725–730.

Corps, R.E., Brooke, C., and Pickering, M.J. (2022). Prediction involves two stages: Evidence from visual-world eye-tracking. *Journal of Memory and Language, 122*, 104298.

Epley, N., Keysar, B., van Boven, L., et al. (2004). Perspective taking as egocentric anchoring and adjustment. *Journal of Personality and Social Psychology, 87*, 327–339.

Epley, N., Morewedge, C.K., and Keysar, B. (2004). Perspective taking in children and adults: Equivalent egocentrism but differential correction. *Journal of Experimental Social Psychology, 40*, 760–768.

Fay, N., Garrod, S., and Carletta, J. (2000). Group discussion as interactive dialogue or as serial monologue. *Psychological Science, 11*, 481–486.

Ferreira, V.S., and Dell, G.S. (2000). Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology, 40*, 296–340.

Ferreira, V.S., Slevc, L.R., and Rogers, E.S. (2005). How do speakers avoid ambiguous linguistic expressions? *Cognition, 96*, 263–284.

Fox Tree, J.E. (2001). Listeners' uses of um and uh in speech comprehension. *Memory & Cognition, 29*, 320–326.

Fox Tree, J.E., and Clark, H.H. (1997). Pronouncing "the" as "thee" to signal problems in speaking. *Cognition, 62*, 151–167.

Garrod, S., and Pickering, M.J. (2004). Why is conversation so easy? *Trends in Cognitive Sciences, 18*, 8–11.

Gebru, T., and Mitchell, M. (2022). We warned Google that people might believe AI was sentient: Now it's happening. *Washington Post* (June 17).

Grice, P. (1989). *Studies in the Way of Words*. Harvard University Press.

Hanna, J.E., and Tanenhaus, M.K. (2004). Pragmatic effects on reference resolution in a collaborative task: Evidence from eye movements. *Cognitive Science, 28*, 105–115.

Hanna, J.E., Tanenhaus, M.K., and Trueswell, J.C. (2003). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language, 49*, 43–61.

Hawkins, R.D., Gweon, H., and Goodman, N.D. (2021). The division of labor in communication: Speakers help listeners account for asymmetries in visual perspective. *Cognitive Science, 45*(3), e12926.

Haywood, S.L., Pickering, M.J., and Branigan, H.P. (2005). Do speakers avoid ambiguities during dialogue? *Psychological Science, 16*, 362–366.

Horton, W.S., and Gerrig, R.J. (2005a). The impact of memory demands on audience design during language production. *Cognition, 96*, 127–142.

Horton, W.S., and Gerrig, R.J. (2005b). Conversational common ground and memory processes in language production. *Discourse Processes, 40*, 1–35.

Horton, W.S., and Spieler, D.H. (2007). Age-related differences in communication and audience design. *Psychology and Aging, 22*, 281–290.

Keysar, B., Barr, D.J., Balin, J.A., et al. (1998). Definite reference and mutual knowledge: Process models of common ground in comprehension. *Journal of Memory and Language*, 39, 1–20.

Keysar, B., Barr, D.J., Balin, J.A., et al. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science, 11*, 32–38.

Keysar, B., and Henly, A.S. (2002). Speakers' overestimation of their effectiveness. *Psychological Science, 13*, 207–212.

Keysar, B., Lin, S., and Barr, D.J. (2003). Limits on theory of mind use in adults. *Cognition, 89*, 25–41.

Kirk, D. (2002). New standards mean Korean Air is coming off many "shun" lists. *New York Times* (March 26).

Kraljic, T., and Brennan, S.E. (2005). Prosodic disambiguation of syntactic structure: For the speaker or for the addressee? *Cognitive Psychology, 50*, 194–231.

Kronmüller, E., and Guerra, E. (2020). Processing speaker-specific information in two stages during the interpretation of referential precedents. *Frontiers in Psychology, 11*, 552368.

Lau, B.K.Y., Geipel, J., Wu, Y., et al. (2022). The extreme illusion of understanding. *Journal of Experimental Psychology: General, 151*(11), 2957–2962.

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

Lockridge, C.B., and Brennan, S.E. (2002). Addressees' needs influence speakers' early syntactic choices. *Psychonomic Bulletin & Review, 9*, 550–557.

Marcus, G. (2022). *Nonsense on Stilts*. https://garymarcus.substack.com/p/nonsense-on-stilts (June 12).

Metzing, C., and Brennan, S.E. (2003). When conceptual pacts are broken: Partner-specific effects on the comprehension of referring expressions. *Journal of Memory and Language, 49*, 201–213.

Nadig, A.S., and Sedivy, J.C. (2002). Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychological Science, 13*, 329–336.

Pickering, M.J., and Garrod, S.C. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27*, 169–226.

Python, M. (1990). *Just the Words: Vol. 2*. Methuen, Mandarin.

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

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

Shintel, H., and Keysar, B. (2007). You said it before and you'll say it again: Expectations of consistency in communication. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 357–369.

Stivers, T., Enfield, N.J., Brown, P., et al. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences, 106*, 10587–10592.

Tanenhaus, M.K., Spivey-Knowlton, M.J., Eberhard, K., et al. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*, 1632–1634.

Vanlangendonck, F., Willems, R.M., and Hagoort, P. (2018). Taking common ground into account: Specifying the role of the mentalizing network in communicative language production. *PLoS One, 13*(10), e0202943.

Wardlow-Lane, L., and Ferreira, V S. (2008). Speaker-external versus speaker-internal forces on utterance form: Do cognitive demands override threats to referential success? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 1466–1481.

Wardlow-Lane, L., Groisman, M., and Ferreira, V.S. (2006). Don't talk about pink elephants! *Psychological Science, 17*, 273–277.

Whalen, J., Zimmerman, D.H., and Whalen, M.R. (1988). When words fail: A single case analysis. *Social Problems, 35*, 335–362.

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

Wu, S., and Keysar, B. (2007). The effect of culture on perspective taking. *Psychological Science, 18*, 600–606.