1. INTRODUCTION

Many of the words used to talk about language use – e.g., communication, coordination, convention, conversation, cooperation, collaboration, community – come from the Latin stems co- or con-, which signify “with” and “together”, respectively (Kennedy, 1890/1996). This confirms that our Latin-speaking predecessors thought of language as part of a social activity that people do “together with” other people. Regrettably, throughout the history of modern psycholinguistics, this appreciation for the social dimension of language has often been absent. Instead of studying how people achieve shared understanding in conversation, psycholinguists have traditionally studied the cognitive structures and processes that underlie the production or comprehension of words or sentences in social isolation. Although a worthwhile enterprise, the development of ecologically valid theories requires greater attention to the cognitive activities involved in actual conversation.

While psycholinguistic research has regarded language as a cognitive skill, language use research has focused on the collaborative dimensions of conversation, often ignoring the cognitive basis of speaking and understanding. This tends to result in theories based on descriptive principles rather than explanatory mechanisms. Fortunately, these two traditions are starting to merge, auguring the development of a theory of language processing that is explanatory, cognitively plausible, and ecologically valid. In this chapter, we review research that lays the groundwork for such a theory.

1.1. Overview and Scope

Language use requires coordination to be successful, because language can be ambiguous. Thus, speakers and listeners will often have to adapt their behavior to their interlocutor in order to avoid misunderstanding. A central question of this chapter is how such adaptations are achieved during the production and comprehension of language.
The specific hypothesis that we investigate in this chapter is the *audience design hypothesis*, or the *design hypothesis* for short. This hypothesis assumes that speakers and listeners achieve success in communication because they maintain detailed models of what the other person knows, and speak and understand against these models (Clark & Murphy, 1982). Given the quantity of evidence currently available, it seems undeniable that speakers adapt some aspects of their speech to characteristics of their listeners. Adults often speak to small children using “motherese,” a form of speech containing exaggerated prosody, a higher pitch, and simplified syntax (Snow, 1972). Speakers change their speaking register or style depending upon the social identity of their addressees (Bell, 1984). Casual observation reveals that bilinguals will often mix languages when talking to other bilinguals, but tend to use a single language when speaking to monolinguals. And developmental psychologists such as Piaget (1926/1955) have observed that the speech of children becomes less egocentric and more listener-centered as they mature (see also Flavell, Botkin, & Fry, 1968).

There is extensive evidence from social psychology that people’s speech varies according to their assumptions about the recipient (for a review, see Krauss & Fussell, 1991). When speakers describe an abstract figure so that another person will be able to identify it, they use longer descriptions that are easier for others to understand than the descriptions that they produce for themselves (e.g., Danks, 1970; Fussell & Krauss, 1989a; Krauss, Vivekananthan, & Weinheimer, 1968; Werner & Kaplan, 1963). When they are asked to describe a figure for a specific friend, the friend is slightly better at interpreting that description than a stranger for whom the message was not intended (Fussell & Krauss, 1989b). Speakers also have some ability to convey a message to an intended audience while concealing it from overhearers (Clark & Schaefer, 1987; Fleming & Darley, 1991). Finally, it has been noted that the lower the perceived identifiability of a stimulus among the members of a social group, the more information speakers will provide when describing that figure to a member of that group (Fussell & Krauss, 1992). All of these findings indicate that speakers take characteristics of the listener into account when they speak.

Therefore, the question is no longer whether speakers design their speech for their listeners, but how and when they do so. In this review, we focus on research that addresses the following questions:

- How fully integrated into the language-processing system is perspective taking? Which levels of processing does it constrain, and which not?
- Does audience design figure into the initial design of a speaker’s utterance, or do its effects arise through monitoring?
- How do listeners exploit the assumption of audience design when they comprehend language?
- Can mechanisms other than perspective-taking mimic the effects of audience design?

Our review begins by introducing the notion of *common ground*, a form of knowledge that has been assumed to provide the basis for audience design. Then, we provide
background for our review in terms of a set of theoretical and methodological prelimi-
naries. The review itself is divided into three main sections: one on language production,
one on language comprehension, and one on repeated reference. We conclude by dis-
cussing the extent to which the body of evidence supports various models for the use of
common ground in language processing.

1.2. Establishing Common Ground

Sometime in your life, you might find yourself seated in a plane next to a guy who you
recognize as member of the same secret society as you are. As joint members of the society,
you would both have extensive shared knowledge: the secret handshake, the mission and
philosophy of your group, the time and location of your annual meeting, the identity of the
society’s president, and so on. However, unless you knew that your companion also recog-
nized you as a member of this secret community, there would be no reason for you to use
this knowledge in interpreting what he says. Thus, when your companion starts a conver-
sation by saying I hear that the president will give a speech today, you might misunder-
stand him if you interpret the president to mean the president of your secret society.
Because you lack evidence that he knows that you are also a member of the same secret so-
ciety, the knowledge is not mutual. As a consequence, you might be better off assuming that
he is referring to the president of the country you are both citizens of.

Clark and colleagues (e.g., Clark, 1992; Clark & Marshall, 1981) have proposed that
a special kind of mutual knowledge, or common ground, provides the critical background
against which speakers produce utterances and listeners comprehend them. In this frame-
work, speakers strive to optimally design their speech with respect to the common
ground: “The speaker designs his utterance in such a way that he has good reason to be-
lieve that the addressees can readily and uniquely compute what he meant on the basis of
the utterance along with the rest of their common ground” (Clark, Schreuder, & Buttrick,
1983, p. 246). Likewise, it is assumed that listeners make the assumption that speakers’
utterances are optimally designed, and use this assumption to guide how they interpret
the speech (Clark & Carlson, 1981; Clark et al., 1983). Although mutual knowledge per
se may never be attainable in practice, common ground can be inferred based on the fol-
lowing types of evidence (Clark & Marshall, 1981):

Physical copresence. Information that is physically copresent is perceptually available
to both interlocutors, who can also perceive that they can both perceive it. Physical or
perceptual copresence is regarded as the strongest evidence for common ground. The
standard example is of a candle that is placed on a table between two interlocutors. Both
can see the candle, and both can also see that they both see it.

Linguistic copresence. Things that are said during the course of a conversation become
part of a “discourse record” that is a subset of interlocutors’ common ground. Thus, if I
tell you Yesterday, I met a man who is an Amway salesman, I can later refer to him using
a simpler description such as the Amway guy or simply the pronoun he, and you can
retrieve the appropriate referent from the discourse record.
Community membership. As in the “secret society” example, interlocutors can infer common ground based on their joint membership in certain communities, once there is something to make this joint membership mutually evident.

1.3. Using Common Ground during Language Processing

A theory of language use in dialogue must describe not only how interlocutors establish common ground, but also how that knowledge is used in processing language. Here are four main contending models:

Full constraint. This model adopts the strongest stance regarding the use of common ground in language processing; thus, it lies at one extreme of the spectrum of possible theories. Under full constraint, all levels of language production or comprehension have full access to common ground, and common ground imposes an immediate, obligatory, and complete constraint on the operation of each level. This account predicts that information that is available to the speaker or listener but that is not part of their common ground would have no impact on language processing.

Partial constraint. This model assumes that common ground does not fully constrain processing, but is one of many cues integrated in parallel during production or comprehension (Hanna, Tanenhaus, & Trueswell, 2003; Nadig & Sedivy, 2002). The probability that common ground will be used will depend on its salience and cue validity relative to other cues. Although the constraint of common ground on processing is partial rather than full, the constraint applies to all levels of production or comprehension from the earliest moments of processing. Partial constraint predicts that private information will have some effect on processing; however, all other things being equal, it should have less of an effect than information in common ground.

Multiple-systems accounts. Multiple-systems accounts assume that the use of common ground is a controlled, optional process, and that it is therefore subject to restrictions on cognitive capacity. The two main existing multiple-systems accounts make different assumptions about how common ground is integrated into language production and comprehension.

One of these accounts, the perspective adjustment model of Keysar and colleagues (Epley, Keysar, van Boven, & Gilovich, 2004; Horton & Keysar, 1996; Keysar, Barr, Balin, & Brauner, 2000; Keysar, Barr, Balin, & Pack, 1998) assumes that common ground is a functionally distinct process that belongs to an “adjustment” stage of processing, but that it imposes no constraint on production or comprehension processes per se. Instead, it assumes that these core stages of language processing are conducted egocentrically, that is, by using information available to the self regardless of whether it is part of common ground. The adjustment stage, which is optional, and slow to take effect, uses common ground to detect and correct violations of common ground. For production, perspective adjustment assumes that utterances are initially designed egocentrically, and
common ground is used to bring utterances into line with the listener’s perspective. For comprehension, it contends that listeners initially interpret utterances egocentrically and only later monitor for, and correct, violations of common ground.

Another multiple-systems account is what we call the dual-process model, formulated for language production by Bard et al. (2000) and Rossnagel (2000). Instead of assuming that the core set of language processes operate egocentrically with a common-ground-sensitive monitor, these accounts make a distinction between automatic and controlled processes in the language-processing system. The model assumes that some language processes operate automatically in a manner that is insensitive to common ground, while others are controlled, and therefore, can be influenced by common ground, given sufficient cognitive resources. For instance, Bard and Aylett (2005) argue that whether a speaker articulates a word more or less intelligibly is determined by an automatic process (priming), while the decision of whether to mark a reference using the indefinite article *a* versus the definite article *the* is a controlled process that is constrained by common ground.

Ordinary memory view. What we call the ordinary memory view (Gerrig & McKoon, 1998; Horton & Gerrig, 2005) is less a theory about the use of common ground than a denial of the psychological reality of common ground. In this model, the effects of common ground are fully reduced to everyday memory processes. As Horton and Gerrig (2005) state: “In general, we wish to ‘normalize’ audience design as a straightforward consequence of how partner-specific information is encoded and retrieved during routine utterance planning ... people will perform utterances that show evidence of audience design whenever the memory representations that permit audience design become accessible with the appropriate time course.” (p. 129). Thus, this theory assumes that knowledge of mutual information is not “consulted” during conversation, but instead, successful coordination emerges naturally from low-level memory processes.

2. THEORETICAL AND METHODOLOGICAL PRELIMINARIES

For an explanatory theory of audience design, it is essential to view the problem of coordinating mutual understanding at three levels of analysis: the cultural, the interactional, and the cognitive levels (Barr & Keysar, 2005). We suggest that this three-level framework is important because it highlights the possibility of multiple explanations for the same phenomenon. In this section, we define these three levels and discuss their implications for the study of audience design.

The cultural level. Successful communication depends on the existence of overlapping cognitive representations among a community of language users. This overlap exists because of the evolution and diffusion of conventions. Conventions can come into existence as a by-product of language users’ efforts to achieve mutual understanding with specific partners (Barr, 2004; Garrod & Doherty, 1994; Hutchins & Hazlehurst, 1995; Steels, 1996). This is consistent with findings that suggest that meaning coordination
during conversation result in enduring changes to how people represent and/or talk about referents (Garrod & Doherty, 1994; Malt & Sloman, 2004; Markman & Makin, 1998). Social network theory suggests that even very large communities can be organized in the form of “small worlds,” where the paths between any two arbitrary individuals can be traversed in a small number of steps (Travers & Milgram, 1969; Watts, 1999). Such short paths enable the rapid diffusion of information throughout the community.

_The interactional level._ Although sharing representations goes some way toward reducing ambiguity, language users will rarely find perfect overlap in their cognitive representations. Thus, interlocutors will often find it necessary to adapt prevailing linguistic conventions to suit their current needs. At the interactional level of analysis this adaptation is achieved via collaborative processes. This may include consulting a model of the interlocutor’s knowledge, monitoring feedback from one’s interlocutor, and making adjustments in response to this feedback. The collaborative model of Clark and colleagues (Clark, 1992, 1996) provides an elaborate theory of the interactional level. Face-to-face conversations provide multiple channels of communication beyond words – including gesture, eye gaze, and prosody – that can reduce the ambiguity of speech. These channels form part of a “collateral track” of communication through which listeners can provide concurrent evidence of understanding without disrupting the current speaker’s turn at talk (Clark, 1996).

_The cognitive level._ The cognitive level of analysis focuses on the mental machinery that underlies language processing in conversation, which includes processes specific to language (e.g., grammatical and phonological encoding), as well as more general cognitive processes (e.g., memory, attention, decision-making). Interlocutors’ ability to successfully communicate will be constrained by the nature and limitations of these underlying processes.

We have highlighted this three-tiered framework because it implies that there will typically be more than one explanation for the behavior of speakers and listeners. Explanations couched at the cultural level will invoke the existence of overlapping representations, dominant language practices, frequency of usage, and so on. They may also invoke language users’ mental models of the capabilities and preferences of the typical community member. Explanations at the interactional level will invoke collaboration, feedback, and the use of common ground. Finally, explanations at the cognitive level will invoke cognitive availability, access to semantic or episodic representations, priming, attention, cognitive capacity, and so on.

The existence of multiple explanations makes language use in dialogue a notoriously difficult subject to study. A good example of how more than one explanation could be provided for the same phenomenon can be seen by considering structural repetition in discourse (Levelt & Kelter, 1982). When speakers say something that follows the same syntax as that of a previous speaker, they may do this because: (1) it is the dominant way to say this in their language (cultural level); (2) they wish to reduce ambiguity for the listener or express agreement with the previous speaker (interactional level); or (3) because
structural priming simply made this particular structure temporarily more available to the production system (cognitive level). Researchers tend to give priority to certain levels of explanation over others, consistent with their pretheoretical commitments. Thus, a psycholinguist would be most likely to explain this phenomenon in terms of structural priming, whereas a researcher working in a collaborative framework would view it as a way of expressing agreement or reducing ambiguity for one’s interlocutor.

In the study of audience design, priority has been given to the interactional level. This has led to the following misconception: if a speaker produces an utterance that is consistent with the common ground with the listener, then that speaker must have consulted common ground in the course of designing the utterance. That speakers modulate their speech depending on context, and that these modulations tend to make their speech easy for listeners to understand, would seem to provide excellent evidence for audience design. However, changes in common ground also change the information available to the speaker (Keysar, 1997). For example, speakers might use the description circle to refer to the round object in a display containing one circle and one square, but would tend to call it small circle in a display containing a larger circle (Olson, 1970). Such behavior is in line with what one would expect from a cooperative speaker (Grice, 1975). However, it is also consistent with the idea that people categorize objects differently in different contexts. Even without any intention to speak, a person looking at such displays may mentally categorize the object as a circle in one display and as a small circle in the other display.

Dell and Brown (1991) were the first to observe that audience design is not the only explanation for speakers’ adaptations. They distinguished between two kinds of adaptations: generic-listener adaptations and particular-listener adaptations. Generic-listener adaptations would benefit comprehension for a generic listener. These adaptations can be made by consulting a model of the generic listener in the language community, or could simply be a by-product of parallelism between language production and language comprehension. In contrast, particular-listener adaptations are based on common ground with the listener.

Dell and Brown illustrate this with an example of the choice of the word lie versus prevaricate to refer to the telling of an untruth. Speakers might choose lie over prevaricate because they estimate that their listener would be unlikely to know the meaning of prevaricate; this choice would reflect a particular-listener adaptation. However, speakers use the word lie over prevaricate simply because it is more accessible to them due to its greater frequency of usage. Precisely because lie is more frequent, it would also be easier for the listener to comprehend. Thus, this choice would be a generic-listener adaptation; specifically, in this case it would be a by-product of how the lexicon is organized rather than the outcome of an assessment of the listener’s knowledge.

Given the possibility of multiple explanations for a given phenomenon, it is necessary to deconfound effects that are due to the speaker’s perspective from the effects that are due to the speaker’s beliefs about the listener’s perspective (Keysar, 1997). To achieve
such separation of perspectives, at least two strategies are appropriate: (1) manipulate the speaker’s perspective while holding constant the speaker’s beliefs about common ground; (2) manipulate the speaker’s beliefs about common ground while holding constant the speaker’s perspective. The first of these two options is often accomplished through the introduction of “private” speaker knowledge. The second option can be implemented by manipulating a speaker’s beliefs about what the listener knows. However, it could also inadvertently affect the availability of information to the speaker (Keysar, 1997).

A final caution in the study of conversational coordination involves a distinction between self-prompted and other-prompted adaptations. Self-prompted adaptations are those that speakers generate spontaneously, even without the addressee’s intervention. In contrast, other-prompted adaptations are due to the intervention of the addressee. Sometimes in experiments on language production, speakers converse with a naïve participant. While this is intended to safeguard ecological validity, uncontrolled listener feedback can cloud interpretation by making it difficult to distinguish between self- and other-prompted adaptations. The only way to distinguish between them is to control for the behavior of the addressee, either by using a trained confederate (Hanna et al., 2003; Keysar et al., 2000; Metzing & Brennan, 2003), a “virtual partner” (Barr & Keysar, 2002), or an imaginary partner (Schober, 1993). Experiments that lack such a control might gain ecological validity, but cannot inform us about the nature of speaker adaptations.

Although the above concerns apply to language production, they are also relevant to language comprehension. Listeners may interpret utterances in line with common ground not because they consulted their common ground with the speaker, but because they used available information that happened to be in common ground. Thus, it is important to design comprehension experiments that either manipulate the information available to the listener while holding common ground constant, or manipulate the common ground while holding the information available to the listener constant. Furthermore, listeners’ adaptations may be self-generated or prompted by speaker feedback. The remedies suggested above also apply to comprehension studies.

3. CURRENT STATUS OF THE DESIGN HYPOTHESIS IN LANGUAGE PRODUCTION

Speakers have many opportunities to take a listener’s needs into account – when they decide what to say, when they formulate their linguistic utterance that they will produce, and when they monitor their utterance plans or their overt speech. In this section, we discuss the evidence for and against audience design at each of these stages.

3.1. Determining What to Say

When listeners comprehend narratives they fill in missing information by accessing long-term memory structures such as scripts, schemata, or frames (Minsky, 1974; Schank
& Abelson, 1977). For example, in comprehending a passage describing a stabbing incident, comprehenders might draw the inference that a knife was used. A study by Brown and Dell (1987) examined whether speakers provide information precisely when the comprehender is less likely to infer it from background knowledge. In their study, participants read and then retold to a confederate, short narratives that described a character performing some action (e.g., stabbing). The action involved an instrument that was either typical (e.g., knife) or atypical (e.g., icepick) for that action. Brown and Dell found that speakers were about twice as likely to mention an atypical than a typical instrument in their retellings, suggesting a possible listener adaptation.

However, the results could also be a by-product of how lexical knowledge is represented and accessed during production. For example, in accessing the verb *stab*, the production process might note that the instrument, an icepick, conflicted with the instrument conventionally used for this action, and therefore might tag the instrument to be explicitly mentioned. Supporting this idea, the majority of explicit mentions of atypical instruments (52%) were linguistically packaged in the same syntactic clause as the verb, in an immediately following prepositional phrase (e.g., *he stabbed the man with an icepick*). In contrast, only 33% of the explicit mentions of typical instruments were in this location. The explicit mention of atypical instruments in the verb clause was not sensitive to the needs of the listener: even when the listeners had already been informed about the instrument, speakers mentioned the atypical instrument at the same rate. However, whether or not the listener was informed did have an impact on the likelihood of explicit mention in a separate clause after the verb (e.g., *The robber stabbed the man. He used an icepick*). Notably, when the listener was uninformed about the instrument’s identity, speakers were simply more likely to mention *any* instrument – typical or atypical – in this location, which might be considered a relatively coarse particularized adaptation. Because the effects occur late in the sentence (in a separate clause after the verb), they are likely to result from a monitoring process – that is, as a kind of afterthought that is appended to the original preverbal message and which did not figure into the initial design of the utterance (Brown & Dell, 1987).

Lockridge and Brennan (2002) challenged this conclusion, arguing that the informational needs of actual addressees would have been different from those of the confederates in Brown and Dell (1987). The confederates, who had heard the stories many times over, may have provided feedback that indicated that they were already knowledgeable about the narratives. Indeed, Lockridge and Brennan found that when naïve listeners lacked pictures, speakers were about 15% more likely to mention atypical than typical instruments within the verb clause. This suggests that speakers are able to make use of listener knowledge in their production of speech. However, it is still possible that explicit mention of the atypical instrument was not included in the original design of the utterance, but was edited into the verb clause by the monitor. Given that speakers and addressees interacted freely, speakers would have had repeated opportunities to learn about their addressees’ specific needs, and may have begun to monitor their speech more carefully before articulation. These results might be reconciled with those of Brown and Dell by assuming that the mention of atypical instruments is a spontaneous, generic
adaptation, but may also constitute a particularized adaptation when the speaker receives adequate listener prompting.

This possibility seems viable given evidence from Horton and Keysar (1996) that certain adaptations may be accomplished via monitoring and adjustment rather than through incorporation into the initial utterance plan. In their study, speakers described a target object to a listener. The target object (e.g., circle) appeared in the presence of a “context” object (e.g., a larger circle) that was either shared with the listener or private to the speaker. Adapting to the listener’s needs, speakers’ target object descriptions were more likely to include adjectives related to the context object (e.g., small circle) when the context object was shared compared to when it was private (75% versus 24%). However, when placed under pressure to begin their descriptions quickly, speakers showed no such sensitivity to the listener’s perspective. Assuming that time pressure obstructs later, monitoring-type processes, this result demonstrates that information about the interlocutor’s perspective is not taken into account during the design of an utterance, but during an additional monitoring stage. (See Polichak & Gerrig, 1998, for a dissenting view, and Keysar & Horton, 1998, for a rebuttal.)

Rossnagel (2000) argued that such adaptation in speaking might be the outcome of both automatic and controlled processes. Automatic processes of message planning make use of cognitively available information, while controlled processes allow for the tailoring of the plan to the addressee’s perspective. As with perspective adjustment, this automatic-controlled account predicts that utterances will be less adapted to the listener’s needs when the speaker is under cognitive load. In Rossnagel’s study, speakers described the assembly of a machine model to an adult university student or to a seven-year old boy, whose limited knowledge would require special tailoring. It was found that cognitive load impaired speakers’ ability to take the addressee’s perspective into account. While speakers varied their instructions depending upon the identity of the addressee in the low-load condition, they gave similar instructions to the student and to the young boy under high load. These results complement those of Horton and Keysar (1996) in support of the monitoring and adjustment hypothesis.

3.2. Avoiding Syntactic Ambiguity

A related question is whether or not speakers take steps to make their speech easier for comprehenders to parse. Speakers can avoid syntactic ambiguity by using optional words, by ordering the constituents in a sentence, or by marking syntactic breaks using prosody.

3.2.1. Structural disambiguation

Most people would initially experience difficulty in parsing the string the coach knew you ate upon arriving at the word ate. This is because the comprehension system prefers to interpret you as the direct object of the verb knew rather than as the subject.
of an embedded sentence. Speakers could circumvent such “garden path” difficulties by including optional words such as the complementizer that, as in the coach knew that you ate. Ferreira and Dell (2000) investigated whether the inclusion of such optional words reflected audience design. They contrasted sentences where insertion of the word “that” would prevent an ambiguity (e.g., the coach knew you ...) with sentences where there was no ambiguity to prevent (e.g., the coach knew I ...). Ferreira and Dell found that speakers are just as likely to include optional words when it would prevent ambiguity as when it would serve no such function (for a similar result, see Kraljic & Brennan, 2005). They found that speakers were only more likely to use optional words when material in the embedded clause was less readily available in memory, a speaker-oriented factor that is independent of the listener’s needs.

The ordering of syntactic constituents is another device that speakers could use to avoid ambiguity. Arnold, Wasow, Asudeh, and Alrenga (2004) investigated whether speakers would use constituent ordering to disambiguate how a prepositional phrase should be attached to the syntactic structure. Such attachment ambiguities arise in sentences containing more than one prepositional phrase, e.g., The judge sent the letter to the president to the committee. Only at the end of this sentence does it become clear that the letter was not sent to the president, but to the committee. Speakers could help listeners avoid such garden paths by positioning the goal argument directly after the verb, as in, The judge sent to the committee the letter to the president. Arnold et al. found that speakers did not use reordering to avoid syntactic ambiguity: they were not more likely to include the goal early in sentences that would otherwise be ambiguous compared to in control sentences that included no such syntactic ambiguity. However, Arnold et al. did find effects of the speaker-oriented factors of syntactic weight and lexical bias on constituent ordering.

3.2.2. Prosodic disambiguation

Speakers can also use prosody to disambiguate syntax, but this does not appear to be a particularized adaptation. Allbritton, McKoon, and Ratcliff (1996) had speakers read aloud syntactically ambiguous sentences that were disambiguated for them by context. They found that speakers – even professionals such as actors and broadcasters – did not reliably use prosody to disambiguate the sentence. Speakers did so only with heavy prompting, such as when they viewed the two different interpretations side by side (see also Fox Tree & Meijer, 2000). However, it is possible that reading out loud is fundamentally different from speaking. To test this, Schafer, Speer, Warren, and White (2000) had participants play a collaborative game in which they used prescripted sentences to communicate about the locations of game pieces. These sentences had a temporary syntactic ambiguity following the main verb, e.g., When that moves the square will ... versus When that moves the square it .... In contrast to Allbritton et al., they found that speakers did prosodically disambiguate the utterances. Curiously, the speakers used disambiguating prosody even when the context already disambiguated the meaning. Therefore, this could not be a listener-oriented adaptation.
Only one study found that speakers might spontaneously use prosody to disambiguate syntax. Snedeker and Trueswell (2003) had speakers play a game with a listener involving actions on a set of objects. The experimenter demonstrated an action for the speaker, and the speaker’s task was to read a sentence from a card to lead the listener, who did not witness the action, to perform the same action on the objects. The utterances were syntactically ambiguous (e.g., *Tap the frog with the flower*) in that the prepositional phrase could be taken as specifying the referent (e.g., which frog to be tapped), or the instrument to be used in the action. When the speaker’s referential context supported both interpretations, speakers provided disambiguating prosody; however, they did not do so when the context allowed only one interpretation (e.g., only one frog). These findings support the idea that speakers only produce prosodic disambiguation when they are aware of an ambiguity, and other disambiguating cues are absent (see also Straub, 1997).

However, Kraljic and Brennan (2005; see also Schafer, Speer, & Warren, 2005) could not replicate this result. They found that the likelihood of prosodic disambiguation depended only on the syntactic structure, and did not depend on the speaker’s perception of situational ambiguity. Kraljic and Brennan did not prescript speakers’ utterances, but used a task that would naturally induce utterances containing prepositional phrase-attachment ambiguities, such as *put the dog in the basket on the star*. Speakers disambiguated sentences using prosodic cues, but they did so even when the referential context already disambiguated them. Furthermore, interacting speakers were no more likely to produce prosodic disambiguation than speakers who spoke alone. These findings show that speakers use prosody for syntactic disambiguation, but not with a listener’s needs in mind.

Another prosodic dimension of speech that might appear to be related to audience design is the clarity with which speakers pronounce individual words in ongoing speech. According to the design hypothesis, speakers should modulate their articulatory effort depending upon the predictability of the word in context (Lindblom, 1990). When a word is not predictable, speakers should expend extra effort to pronounce it clearly, because the listener would have to rely mainly on bottom-up information to identify the word. In support of this view, listeners tend to articulate the word *nine* more clearly in the sentence *The next number you will hear is nine* than in *A stitch in time saves nine* (Lieberman, 1963). In addition, words used to refer to “given” entities in discourse tend to be shorter in duration than words used to refer to “new” entities, with vowels often taking on a phonologically reduced form, e.g., *[i]* to *[ɘ]*, (Bard et al., 2000; Fowler & Housum, 1987). At first blush, such findings seem to be evidence for audience design, because they seem well-adapted to listeners’ perceptual needs. However, Bard et al. (2000) found that speakers pronounced words less clearly whenever they were repeated and coreferential, regardless of whether they were given or new for the listener. This result suggests that articulatory reduction is the result of speaker-centered priming processes, and is not a true listener adaptation.

### 3.3. Marking Referent Accessibility

Interestingly, one area in which researchers have found evidence for audience design is in the speaker’s marking of the accessibility of referents (Bard & Aylett, 2005; Hupet &
Chantraine, 1992; Lockridge & Brennan, 2002). Linguistic theory suggests that speakers mark the accessibility of referents using variations in referential form (Ariel, 1988; Chafe, 1976; Gundel, Hedberg, & Zacharski, 1993). New referents – referents not yet introduced into the discourse – are generally of low accessibility. Speakers tend to mark new referents using the indefinite article a accompanied by an elaborate noun phrase (a man I met yesterday on the plane). Referents that have already been introduced (“given” referents), in contrast, tend to be more accessible; accordingly, they are often marked using the definite article the accompanied by a short descriptive phrase (the man). The choice between indefinite versus definite article per se need not indicate a particularized adaptation; speakers might simply mark the accessibility of a referent based on its availability in their own model instead of their beliefs about its accessibility in the listener’s model. However, studies that have independently manipulated the speaker’s and the listener’s knowledge suggest that the choice of definite versus indefinite marking is indeed a particularized adaptation (Bard et al., 2000; Bard & Aylett, 2005; Hupet & Chantraine, 1992).

3.4. Monitoring and Awareness of Ambiguity

Referential ambiguity is non-linguistic in nature: it derives from the mapping between words and the world. In contrast, linguistic ambiguity arises out of linguistic structure and linguistic performance. There are many different sources of linguistic ambiguity that could cause miscomprehension, such as homophony (words that sound alike but that have different denotations, e.g., bat as in flying mammal versus bat as in baseball or cricket bat), ambiguity in phonological segmentation (e.g., alone versus a lone), or syntactic ambiguity (e.g., attachment ambiguities, such as the girl saw the boy on the bike). Furthermore, there are many production variables that can affect the comprehensibility of utterances, such as speech rate, loudness, and clarity of articulation. For these reasons, speakers need to monitor the comprehensibility of what they say.

Ferreira, Slevc, and Rogers (2005) demonstrated that linguistic and non-linguistic ambiguity are monitored by different mechanisms. Speakers described targets such as a vampire bat in the context of a foil object. The foil object was either from the same category (e.g., a larger vampire bat), thereby creating a non-linguistic ambiguity, or from a different category covered by a homophonous term (e.g., a baseball bat), resulting in a linguistic ambiguity (i.e., since the word bat could be used for both types of object). In either case, speakers should avoid the bare homophone and include disambiguating information (e.g., small bat or vampire bat). Although speakers nearly always avoided non-linguistic ambiguity, they rarely avoided linguistic ambiguity except when they had already used the homophone to describe the foil. For example, they were only likely to call the flying mammal a vampire bat when they had just referred to the baseball bat as the bat. This suggests that when speakers monitor their speech before they begin articulation, they will reliably detect non-linguistic ambiguities but not linguistic ambiguities. The difference in the likelihood of detection supports the idea of distinct monitoring mechanisms.

Another mechanism that must be involved in audience design is one that allows the speaker to evaluate how well an utterance is designed for an addressee. Keysar and Henly
(2002) investigated how accurate speakers are in gauging the effectiveness of their utterances. In their study, speakers read aloud syntactically or lexically ambiguous sentences, such as “the man is chasing a woman on a bicycle,” trying to convey one of the meanings to a naïve listener that was seated opposite of them. The pair was then presented with the two paraphrases. Speakers selected which of the two meanings they thought the listener understood, while listeners selected the meaning they thought the speaker had intended. The majority of speakers overestimated the likelihood that their listeners had understood them: they assumed that their meanings were successfully conveyed in 72% of the cases, while listeners actually understood the intended meanings in only 61% of the cases. Although speakers overestimated their effectiveness, overhearers who knew the speakers’ intentions did not. This indicates that the overestimation arises out of the very act of speaking. Such overestimation puts limits on speakers’ ability to engage in audience design.

3.5. Audience Design in the Production of Written Text

Writing differs from speaking in important ways. Unlike authors, speakers produce language under circumstances that leave little time for reflection, leading to greater difficulties in formulating what to say. Speakers also are likely to have immediate feedback from the recipient, which authors lack. Thus, successful communication in writing might demand deeper engagement in perspective taking, but may also offer authors more time to do so.

Traxler and Gernsbacher (1992) demonstrated the difficulty of the lack of feedback for writers by showing that even minimal feedback will improve writers’ effectiveness. They had writers describe a set of abstract figures. One group of writers were later told how many readers were able to identify the correct figure based on their description. This procedure was repeated three times. Writers who received feedback improved the effectiveness of their descriptions, while writers who did not receive feedback showed no such improvement. In a follow-up study, Traxler and Gernsbacher (1993) found that experience in perspective taking – by performing the selection task as a reader – helped writers improve the effectiveness of their messages. Envisioning the addressee’s perspective, then, is a barrier to effective communication, and even minimal feedback can help people communicate more effectively.

4. CURRENT STATUS OF THE DESIGN HYPOTHESIS IN LANGUAGE COMPREHENSION

The audience design hypothesis that was posed for language production can also be posed for comprehension: do people interpret utterances as if they were optimally designed with respect to their common ground with the speaker or writer? Comprehension, like production, also proceeds incrementally through multiple stages, although in the opposite direction: from sound or print into a meaning represented in the mind. The comprehension system must identify words within the perceptual input and then access
syntactic and semantic information from the lexicon in order to build a sentence frame and construct an initial interpretation. Referents are identified within a discourse model or established as new. World-knowledge (scripts and schemata) may need to be consulted to draw inferences and fill in missing information. The ultimate goal of this process is to recover the speaker’s or writer’s communicative intention. Finally – though often overlooked – listeners, like speakers, monitor the coherence and adequacy of the emerging representation (Keysar et al., 1998; Markman, 1977).

At any or all of these stages, the comprehension system might consult a model of the speaker’s knowledge. As was the case for production, the interesting question is not whether common ground constrains comprehension, but at what stage. To date, research has focused on the use of common ground in finding referents for spoken expressions as well as in comprehending written text. In this section, we review these findings and their implications.

4.1. Comprehending Spoken References

Clark et al. (1983) sought to demonstrate that listeners use common ground in interpreting references. They pointed out that demonstrative references as in *that man* are used felicitously even when the term could apply to more than one referent in the context, suggesting that this is so because listeners can use common ground to find a unique referent. For instance, a speaker could point toward two men, a fat and a thin one, and say *George will look like that man very soon*, so long as one of the men is salient with respect to the common ground of the speaker and the listener. If the interlocutors had just been discussing George’s obsession with weight loss, then the listener might take the thin man as the referent. If they had just been discussing his eating binges, then the listener might instead consider the fat man as the referent. In a series of experiments, they found that this is precisely what listeners do. However, listeners in these studies may have responded based on what was salient to them at the moment, rather than on their common ground with the speaker (Keysar, 1997).

A growing number of studies have sought to address listeners’ use of common ground to identify referents using eyetracking. Eyetracking is a powerful technique that provides a moment-by-moment record of the comprehension process (e.g., Cooper, 1974; Eberhard, Spivey-Knowlton, Sedivy, & Tanenhaus, 1995). This makes it possible to observe how listeners progressively narrow the domain of reference as speech unfolds (see chapters in Trueswell & Tanenhaus, 2005, for discussion). Typically, in such experiments listeners take part in a referential communication task in which they follow a speaker’s instructions, which involve looking at or manipulating objects.

In general, eyetracking studies of spoken language comprehension find strong, immediate effects of various kinds of contextual constraint on language processing. For instance, Eberhard, Spivey-Knowlton, Sedivy, and Tanenhaus (1995) found that listeners interpreted complex descriptions containing multiple prenominal modifiers (e.g., *touch the large starred rectangle*) by incrementally constraining the domain of reference as the
speech unfolded. For instance, upon hearing the adjective *large*, they tended to look at objects in the display that contrasted in size (e.g., a large and a small rectangle). Tanenhaus, Spivey-Knowlton, Eberhard, and Sedivy (1995) reported that when the visual context supported a nonpreferred syntactic structure, listeners did not appear to experience a garden path. Similarly, Chambers, Tanenhaus, Eberhard, Filip, and Carlson (2002) found that when listeners interpreted instructions such as *put the cube inside the can*, they restricted attention to cans that would be big enough to accommodate the cube. Finally, Hanna and Tanenhaus (2004) found that when listeners interpreted a speaker’s references, they were attentive to whether the speaker was momentarily able to physically reach various objects.

The above studies suggest that various kinds of contextual constraint can immediately influence how listeners process language. The fact that listeners are able to so flexibly adapt comprehension to context might be viewed as evidence that listeners expect speakers to design their utterances with respect to common ground. For example, a listener might only expect a speaker to refer to a square using a modifier such as *large* if there is more than one square in the listener’s common ground with the speaker. However, none of these studies is designed to permit such an inference, since they do not separate the information available to the listener from the listener’s beliefs about the speaker’s knowledge. Processing language with respect to the available context – whether the context is instantiated in a visual display or in a prevailing discourse model – could be a generic adaptation that would tend to promote successful comprehension whenever information that is available to a listener is also available to the speaker (Barr & Keysar, 2005; Pickering & Garrod, 2004).

The critical question is to what degree is a listener’s search for referents restricted to copresent information. The full constraint model predicts that listeners would only consider referents that are in their common ground with the speaker (Clark and Carlson, 1981). In contrast to the full constraint model, perspective adjustment predicts that listeners would consider referents whether or not they are common. Evidence against full constraint and in favor of perspective adjustment was provided in a series of studies (Keysar et al., 2000, 1998; Keysar, Lin, & Barr, 2003). For example, Keysar et al. (2000) tracked participants’ eyes as they played a communication game with a speaker. In the game, the speaker was a confederate who instructed the participant to move objects around a grid that was placed between them. To keep common ground constant while manipulating the listener’s private knowledge, some objects were mutually visible (i.e., copresent), while others were visible only to the listener (i.e., private). For instance, in one item, the speaker told the listener to *move the small candle* in a display containing two copresent candles that varied in size. While the target was obviously the smaller of the two copresent candles, the listener could also see an even smaller, private candle (the “competitor”). In contrast to the full constraint hypothesis, listeners considered the competitor object as the referent: they were more likely to look at it than at a control object, and the presence of the competitor also delayed the identification of the target object. The most surprising result was predicted only by perspective adjustment: about a quarter of
the time, listeners attempted to move the private competitor, with the majority of listeners attempting this at least once during the experiment.

Such egocentric errors have been found to persist despite extensive attempts to eliminate them. Keysar et al. (2003) had listeners hide the private competitor object from their own view in an opaque brown bag, so that they knew of its existence in the bag but could not see it. Despite this, listeners still considered the private competitor in the bag as the referent. To make it even more obvious that the speaker did not know the identity of the competitor, another experiment led listeners to believe that the speaker had a false belief about the identity of the private competitor in the bag. This manipulation made the experiment similar in logic to the false belief task commonly used with children to assess Theory of Mind (Astington & Gopnik, 1988; Wimmer & Perner, 1983). Surprisingly, listeners showed the same interference when they believed the speaker had a false belief as when they believed she was ignorant. This shows that the cognitive system for reasoning about another’s beliefs is not fully incorporated into the comprehension system, since it appears unable to constrain the search for referents.

Epley, Morewedge, and Keysar (2004) found further evidence for perspective adjustment in a study of perspective taking that directly compared the performance of adults with that of children. The fact that adults are less egocentric than children could be explained in two ways: (1) that adults are better at using mutual knowledge to avoid generating egocentric interpretations; or (2) that egocentric interpretations are generated automatically by the comprehension system, and adults are better at using common ground to monitor and reject these interpretations. Epley et al. found that although adults and children were equally likely to consider the private competitor, adults were faster to recover and were less likely to erroneously select the private object. These results support the idea that language is processed against an automatic egocentric default, and that the main difference between children and adults lies in adults’ ability to detect and correct this egocentric interpretation.

Nadig and Sedivy (2002) and Hanna et al. (2003) argued that the results presented in support of perspective adjustment were also consistent with the partial constraint hypothesis, which assumes that common ground is but one of many cues integrated in parallel by a fully interactive comprehension system. Under this view, the effects of common ground would be immediate but only partial, since other cues might activate information that is available to the listener but is not in common ground. To show that comprehenders are initially egocentric, these researchers suggested, it is necessary to show that they would be equally likely to fixate a private competitor as they would a shared competitor that was an equally good fit to the speaker’s description. Partial constraint would predict that comprehenders would be less likely to fixate a private competitor than a shared competitor during the early moments of comprehension, thereby demonstrating an early, albeit partial, effect of common ground. Perspective adjustment would predict that listeners would initially be equally likely to fixate private and shared competitors.
To test these hypotheses, Nadig and Sedivy (2002) had five-year-old children play a communication game with an adult confederate speaker while their eye movements were monitored. As in Keysar et al. (2000), Nadig and Sedivy found interference from private competitors, but they also found greater facilitation for the target with a private competitor than with a shared competitor, suggesting that common ground imposes a partial constraint on comprehension from the earliest moments. Hanna et al. (2003) found similar results in an eyetracking study with adults. From the earliest moments of comprehension, listeners were more likely to look at a target shape that was in common ground than a matching private shape. This provides further evidence against complete egocentrism, and supports the partial constraint hypothesis.

At present, the literature on common ground use in the comprehension of referential expressions suggests the existence of partial effects of common ground from the earliest moments of comprehension. At the same time, the egocentric element in comprehension has proven surprisingly resilient to manipulations that highlight the difference in perspective between speaker and listener. An important question that has not yet been addressed is whether the early effects of common ground reflect use of common ground by the comprehension system, or a strategic, non-linguistic use of common ground. Listeners may have been less likely to look at private objects than shared objects even in the absence of any linguistic input, as Keysar et al. (2000) found. Thus, copresent and private objects may have different baseline probabilities of fixation at the initial moment of the referring expression. Experiments that involve direct comparisons between shared and private objects do not control for this baseline difference, and thus leave open the possibility that the “partial” effects of common ground are due to task-based strategic effects of common ground mixed with automatic “egocentric” effects.

Although the debate over the use of common ground in the interpretation of spoken references is ongoing, it has already produced substantial insight into the integration of linguistic and contextual information in language comprehension. The use of common ground in non-conversational contexts, such as in text processing, has also increased our understanding of how multiple sources of information are brought together during language comprehension.

4.2. Using Common Ground in Text Processing

4.2.1. Difficulty in using common ground

Effective story comprehension requires readers to keep track of what characters believe and to use this information in interpreting their statements and actions (Graesser, Bowers, Bayen, & Hu, 2001; Graesser, Bowers, Olde, & Pomeroy, 1999). One potential difficulty for readers might be in considering how a given character would interpret the statement of another character, because this requires simultaneously keeping in mind both characters’ points of view. Keysar (1994) investigated such perspective taking with a variety of scenarios. For example, people read that David asked June for a restaurant
recommendation, and that he either had an excellent dinner there (positive experience) or a miserable one (negative experience). He then left her the message “About that restaurant, it was marvelous, just marvelous.” Readers were to assess how June would interpret the message. Given that she was uninformed about the experience, the assessment should have been the same in the two conditions. However, readers were more likely to attribute the perception of sarcasm to June when they knew that David was being sarcastic (negative experience). This phenomenon was termed the “Illusory transparency of intention.”

Weingartner and Klin (2005) showed that the illusion of transparency is not restricted to explicit judgments but occurs during normal reading. They presented readers with the original scenarios that ended with a target sentence that demonstrated that June either understood David’s intention or misunderstood him. Readers were slower to understand the target sentence when it was inconsistent with David’s intention, suggesting that readers indeed inferred that June understood what he meant even though she did not have sufficient information. Keysar (1998) showed that it is the communicative intention of the speaker that appears transparent, and not the speaker’s true attitude: when David’s experience was miserable, but he wanted to conceal this negative information from June, readers no longer thought that June would perceive sarcasm (this result was also replicated in Gerrig, Ohaeri, & Brennan, 2000, and in Weingartner & Klin, 2005).

The phenomenon of the illusory transparency of a speaker’s intention is striking, given that listeners can readily differentiate between what different characters know (Graesser et al., 2001). Thus, it suggests a dissociation between what readers know about what characters know, and how they use that knowledge to assess the perception of intentions. Researchers have offered two explanations for the phenomenon. The first of these explanations assumes that readers, like interlocutors, observe the principle of audience design (Gerrig et al., 2000; Keysar, 1994; see also Keysar, 2000; Gerrig et al., 2000, for discussion). Gerrig et al., (2000) argued that readers will go to great lengths to preserve the notion that speakers are being cooperative. Readers, then, might have assumed that speakers would not have made sarcastic utterances unless they had sufficient reason to believe that the addressee would have some way of perceiving their underlying intention. Thus, readers believe that the “message gets through” because they assume that characters strive to be cooperative, and proceed on their best estimate of the common ground between speaker and listener. Although this view is in line with standard pragmatic theory, there is no direct evidence to support it yet.

The second explanation for the phenomenon assumes perspective adjustment: readers compute the actual intention of the speaker and then try to make allowances for the perspective of the addressee. But as with other anchoring and adjustment phenomena, they adjust insufficiently (Epley et al., 2004; Keysar, 1994). Epley et al. (2004) demonstrated that illusory transparency is exacerbated when people are under conditions that hinder the adjustment process, such as time pressure. The difficulty readers have in taking the perspective of characters, then, might be a special consequence of the difficulty of taking perspective in general.
4.2.2. When is common ground used?

Some studies offer a more optimistic assessment of readers’ effectiveness in taking characters’ common ground into account. Gibbs, Mueller, and Cox (1988) investigated the role of common ground in asking and comprehending questions. Participants read scenarios in which the common ground between characters was manipulated. After each scenario, participants were asked to select the most appropriate question for a speaking character to ask a listening character. There were three alternatives, reflecting different presuppositions about the common ground between the protagonists. The most common selection was in accordance with the common ground that had been established between the speaker and listener. In addition, in a reading comprehension study listeners took less time to read questions that appropriately specified the common ground between the interlocutors. These results indicate that readers are sensitive to common ground in developing expectations about how characters will interact.

The correct interpretation of many kinds of ambiguous utterances depends upon what communities the interlocutors jointly belong to (Clark & Marshall, 1981). For instance, consider how you would interpret a comment regarding a university sports event, such as that game was a disaster. If the speaker is a member of your university, then you would have evidence that your university’s team lost. In contrast, if the speaker is a stranger, then the interpretation would be less certain. Gerrig and Littman (1990) examined how readers exploit protagonists’ community membership in order to interpret their statements. Participants read a set of scenarios, each of which ended with an ambiguous statement. They were asked to interpret such statements as if they were addressed to them. Gerrig and Littman found that readers were more likely to interpret such ambiguous statements in line with community expectations when the speaker was a friend, compared to when the speaker was a stranger.

Greene, Gerrig, McKoon, and Ratcliff (1994) suggested that common ground affects reading by modulating the accessibility of information in memory. They argued that as characters come together, their common ground becomes more accessible in the reader’s mind. In their experiment, participants read scenarios about two characters discussing a third character; one of the two characters then left, and then, after a short interlude, was reunited with the first character. Using a recognition probe task, Greene et al. found that the reunion of the characters could restore the accessibility of information related to the third character, who was in common ground. They suggested that, “upon the reunion of the two characters, the reader makes accessible the common ground that the characters share, in preparation for understanding their future interaction.” (p. 524)

However, Lea, Mason, Albrecht, Birch, and Myers (1998; see also Keysar, 1997) demonstrated that this reunion effect reflects general memory associations rather than the use of common ground per se. Memory-based models (e.g., Gerrig & McKoon, 1998) stipulate that information can be reactivated through a memory “resonance” process that is based on the overlap between concepts. Thus, in the above example, the return of one of the characters may have reactivated concepts associated with her, regardless of...
whether they were part of the common ground between the characters. Indeed, Lea et al. found that the reunion effect extended equally to concepts not in common ground between the protagonists. This supported the idea that the reunion effect was due to memory-based resonance processes, and not to the use of common ground.

In contrast to Greene et al., who argued that common ground modulates memory processes, Gerrig and McKoon (1998) rejected altogether the psychological reality of common ground in language use: "our account denies that common ground is something that speakers (or addressees) can or cannot take into account – merely memory processes acting on representations." (p. 82) and "to the extent that ‘common ground’ seems to work pretty well, we suggest that it is because people serve as highly valid memory cues" (p. 81). However, this view is inconsistent with several observations in the literature. Lea et al. found that though concepts were reactivated independently of common ground, when a protagonist said something that violated common ground with the other, readers were surprised. Clearly, they should not have noticed a violation if they did not take common ground into account. It has also been observed that readers will not use information associated with a speaker when interpreting that speaker’s utterance, if they believe that the speaker wishes to keep that information concealed from the addressee (Gerrig et al., 2000; Keysar, 1998). Finally, it is inconsistent with results from eyetracking studies, which clearly show that making a decision about whether or not to use available information during comprehension involves the use of common ground, whether through monitoring (Keysar et al., 2000) or through some other mechanism that is more integral to the comprehension system (Hanna et al., 2003). Against the background of such a broad range of evidence for the use of common ground in interpreting language, attempts to deny the psychological reality of common ground are not compelling.

5. REPEATED REFERENCE IN DISCOURSE

Patterns of language use tend to evolve over the course of a conversation as interlocutors adapt the conventions of their language to current conversational needs. Most research on such adaptation has focused on how speakers adjust the speech they produce, although there is evidence that listeners adjust their expectations about the speech they are likely to hear (Barr & Keysar, 2002). The majority of these studies focus on repeated reference, i.e., on the changes that occur when speakers repeatedly make reference to an object during conversation. The main phenomena associated with repeated reference are reductions in the number of words used to describe a referent, stability in lexical choice, and changes in articulatory quality. In this section, we review these phenomena and discuss whether they reflect generalized or particularized adaptations.

5.1. Abbreviation of Description

Zipf (1935) noted an inverse relationship between the length of a word and its frequency of use, a relationship known as “Zipf’s Law.” A similar inverse relationship
exists between the length of a description of an object and the number of times the speaker has referred to it within the conversation (Krauss & Weinheimer, 1964, 1966). Typically, when speakers refer to the same referent multiple times, the overall length of their descriptions, as measured in number of words or in speaking turns, declines. Over time, the description takes on “name-like” properties, functioning more as a cue to memory rather than an actual description of the object (Carroll, 1980). In most studies observing this phenomenon, speakers describe abstract figures. The number of words they use is plotted against frequency of use, yielding an exponentially decaying function.

One explanation for this phenomenon is in terms of the accumulation of common ground between participants. This “collaborative model” assumes that interlocutors strive to minimize their collaborative effort in establishing reference through the establishment of mutually accepted descriptions (Clark & Brennan, 1991; Clark & Wilkes-Gibbs, 1986). This view predicts that as common ground accumulates, speakers can shorten their descriptions because addressees will need less information. Clark and Wilkes-Gibbs (1986) observed this process of shortening in a referential communication task where speakers described a set of abstract figures. However, their experiment lacked a control condition to show that speakers who described the figures to different addressees each time would not shorten their descriptions. This left open the possibility that the length reduction is due to speaker-related factors.

Hupet and Chantraine (1992) tested this possibility with two groups of participants performed a referential communication task. Both groups of participants spoke into a tape recorder with no physically present addressee. Participants in one group were told that they would keep talking to the same addressee as they did on the previous trial, while those in the other group were told that they would be talking to a different addressee each time. Hupet and Chantraine found that speakers who spoke to different addressees each time actually increased the number of words they used, while those speaking to the same addressee did not decrease the number of words they used. These results suggest that the presence of addressee feedback may be critical for the reduction in length of referential descriptions (see also Krauss & Weinheimer, 1966). However, in a later study, Hupet and Chantraine (1994) compared groups of participants who described abstract figures to the same addressee in either a monologue or a dialogue, and found that a subset of the participants in the monologue condition behaved the same as the participants in the dialogue condition. This suggests that the reduction effect is only partly due to collaboration.

Isaacs and Clark (1987) looked for evidence for the collaborative model using a referential communication task in which participants sought to communicate the ordering of pictures of New York landmarks. This made it possible to observe when speakers opt for a description of a referent (e.g., the peaked building) versus a proper name (e.g., the Citicorp Building). According to the collaborative model, speakers should use proper names alone only when they have evidence that their listeners are familiar with New York landmarks. A main finding was that experts who spoke to novices often started out using proper names and descriptions, and over time, increased their use of descriptions alone. Novice directors who spoke to expert matchers started out using descriptions and
increased their use of proper names, which they could only have learned from the matcher. These results indicate that, during conversation, speakers and listeners are able to discover their common ground and adjust their language use accordingly. Similar results were reported by Nohara-LeClair (2001) in a series of experiments in which participants made reference to international flags. Participants’ shared knowledge of flag names increased during the referential communication task, as well as their accuracy in their assessments of their partner’s knowledge. In addition, speakers tended to use the name of the country in referring to the flag only when they believed that their partner shared that knowledge; otherwise they tended to use descriptions.

5.2. Conversational Precedents

Another phenomenon of repeated reference is that speakers’ lexical choice tends to stabilize over time. Such “lexical entrainment” (Brennan & Clark, 1996; Garrod & Anderson, 1987) is interesting because it suggests that a speaker’s decision regarding how a referent is to be described depends not only on the current context, but also on the history of usage within a given conversation. When speakers initially select a label for a referent, they have multiple options that vary in their specificity: a certain car can be called car, sportscar, fancy car, vehicle, etc. One factor that will influence a speaker’s initial choice is the set of objects from which the referent is to be distinguished (Olson, 1970). Thus, in the context of a flower, speakers might refer to a car by its basic-level term car. However, in the context of a second car, a speaker would need to choose a more specific term, such as sportscar. Yet, once speakers use a particular term for a referent, they tend to continue to use it on subsequent turns. One important consequence of this is that it can lead speakers to “overspecify” the referent; that is, to refer to a car as the sportscar when there is no other car in the context (Brennan & Clark, 1996). On the surface, such usage would seem to violate Grice’s Maxim of Quantity (Grice, 1975), according to which speakers should provide no more information than necessary. However, such usage could be considered cooperative, in that it exploits established agreements or “precedents” on how a referent is to be conceptualized, or what Brennan and Clark (1996) called “conceptual pacts.”

Brennan and Clark (1996) argued that speakers use conversational precedents in a partner-specific manner; that is, in a manner that is sensitive to common ground. They used a task that induced speakers to entrain on subordinate-level descriptions for pictures of everyday objects (e.g., sportscar instead of car, loafer instead of shoe), and in a subsequent test phase manipulated whether the speakers continued on with the same partner or with a new partner. In the test phase, the pictures appeared in displays lacking same-category objects, so speakers could simply use basic-level descriptions (e.g., car and shoe). The prediction was that speakers would be more likely to do so with a new partner than with an old partner, because the precedents of using subordinate terms are specific to the old partner. Although this prediction was confirmed, in the first trial of the test phase, speakers used subordinate terms with new partners just as much as with old ones. This suggests that the speaker’s abandonment of the precedents with the new addressee may have been other-prompted rather than self-prompted; that is, speakers might have
used the precedents regardless of whether the partner was new or old, but this might have confused the new addressee, prompting speakers to adapt to them.

Horton and Gerrig (2002) provided further evidence for the role of other-prompted adjustments in audience design. They suggested that audience design is only sometimes necessary for references to be successful (which means that listeners will often find speakers’ egocentric descriptions adequate). They hypothesized that the need for audience design will depend on the current conversational task, and that speakers can discover the appropriate circumstances through experience. Horton and Gerrig had participants play a referential communication task with two independent matchers who had different subsets of knowledge. In initial rounds of the game, directors independently established precedents for referring to specific objects with each of the two matchers. There were two later rounds that immediately followed a partner switch, confronting directors with the task of describing referents for which precedents existed with the previous partner, but not the current one. Horton and Gerrig found more evidence of audience design after the second partner switch than after the first one, suggesting that the feedback directors’ received after the first switch prompted awareness of the need to take listeners’ knowledge into account. Thus, speakers did not deploy audience design in an absolute manner, but did learn to detect cases in which it would be necessary.

Bromme and colleagues (Bromme, Jucks, & Runde, 2005; Bromme, Jucks, & Wagner, 2005; Jucks, Bromme, & Becker, 2005; Jucks, Bromme, & Wagner, 2005) investigated audience design where sensitivity to the audience could have serious consequences – in doctor–patient communication. Jucks et al. (2005) investigated entrainment in how medical experts use language when answering e-mail queries about medical problems with the help of a “concept map.” They found that when a query used technical medical terms such as “arteriosclerosis,” the expert was more likely to answer with technical language compared to when the query used everyday language such as “vascular hardening.” Such entrainment could reflect audience design, as the patient who uses technical language is probably more knowledgeable about medicine than the one who does not. But Jucks et al. showed that it actually reflects the availability of the term for the speaker: such entrainment occurred even when the technical term came from the concept map, and not from the patient. Interestingly, though experts’ reflections showed a sensitivity to patient terminology, their answers to the queries did not take advantage of this knowledge. Therefore, even when the stakes are higher than in the harmless activity of identifying cars and candles, speakers use precedents because they are there, not because they are shared with the addressee.

If conversational precedents are used strictly in a partner-specific manner, then speakers should not carry over the precedents from one partner to the next. However, such carry-over could be a cultural-level adaptation that contributes to the diffusion of conventions within a language community. Malt and Sloman (2004) tested whether the precedents established by a speaker in one conversation would be used in a distal conversation. A confederate director used one of two conventional labels (e.g., trash can or waste basket) to describe common objects to a participant matcher in a referential
communication task. The matcher then became a director for a new participant matcher, and the task was repeated. Finally, the new matcher became a director with a new participant. Even though none of the participants in this third conversation interacted with the confederate who originally established the precedent terms, those terms were more likely to show up in this third conversation than the alternate terms. This lends some support to the idea that speakers do not necessarily use precedents in a partner-specific manner.

Barr and Keysar (2002) investigated whether precedents benefit language comprehension, and if so, whether this benefit is partner-specific or partner-independent. They monitored listeners’ eye movements in a referential communication task. Unsurprisingly, listeners were faster to identify referents when referring expressions had precedents, compared to when they were new. This comprehension benefit could be the result of the accumulation of common ground, or could be due to the availability of the precedent. A common ground account would predict that if the second use of the term was made by a new, uninformed speaker, the benefit would be reduced, while the availability account would predict an equal benefit. Barr and Keysar found that listeners benefited equally from precedents, regardless of whether or not the speaker was the one who established the precedent or a new speaker. This result is surprising because the objects used in the experiments were highly unconventional, and therefore it would be highly unlikely for two independent speakers to refer to such objects in precisely the same way.

But perhaps listeners inferred common ground between the two speakers precisely because those speakers choose the same names for an unfamiliar object. For this reason, Barr and Keysar conducted an additional experiment in which a speaker broke a precedent instead of following it. If precedents are used in a partner-specific manner, then the breaking of a precedent should lead to greater interference when it was broken by the same person who established it, as compared to when it was broken by a new speaker. Barr and Keysar found equally strong interference when precedents were violated by a new or old speaker. Thus, listeners appear to use precedents when comprehending speech because they are cognitively available, not because they are part of their common ground with the speaker.

However, in this latter experiment, speakers referred to familiar, everyday objects, and broke precedents by reverting from subordinate-level descriptions (sportscar) to basic-level descriptions (car). Metzing and Brennan (2003) sought to replicate Barr and Keysar’s findings of partner-independence, but used unconventional objects, for which the violation of a precedent might be more jarring. They replicated Barr and Keysar’s finding that listeners were fast to identify referents when precedents were maintained, regardless of speaker. However, they also found that listeners were slower to identify the old-referent target when speakers broke their own precedents than when a new speaker broke a previous speaker’s precedents.

Although the results of Metzing and Brennan do support the idea that listeners use common ground to deal with broken precedents, they leave open at least two important questions. First, since Metzing and Brennan did not provide detailed time-course
information, it is not clear whether their results reflect “early” effects of common ground, as would be predicted by the partial constraint hypothesis, or “late” effects of common ground, as would be predicted by perspective adjustment. Second, the results might reflect a more general expectation that speakers would use terms consistently, independently of their addressees. That would render such results speaker-specific, but without involving common ground.

6. SUMMARY AND PROSPECTUS

We have made the case that the existence of multiple explanations is intrinsic to the study of language use, since it is an activity that can be analyzed at the cultural, interactional, and cognitive levels. Researchers often overlook these multiple explanations due to the entrenchment of theoretical frameworks (whether interactional or psycholinguistic) that give priority to certain levels of explanation over others. We proposed that to properly interpret experimental findings, it is necessary to observe distinctions between generic and particularized adaptations as well as between self-prompted versus other-prompted adaptations. In light of these distinctions, we reviewed evidence for audience design in language production, language comprehension, and repeated reference. In this section, we attempt to assess the degree of support this body of evidence provides for the various models of the use of common ground in language processing.

6.1. Summary of Findings and Evaluation of the Models

When participating in a conversation, language users are beholden to the maxims of cooperative communication (Grice, 1975). Thus, when deciding what one wishes to say to an interlocutor, what kind of speech act to use, what topic to converse about, and what language to use, speakers might extensively consult their knowledge about common ground. By comparing the speech produced by adults to the speech produced by children, one can see that adults pay attention to the interlocutor’s perspective in making such high-level decisions. Researchers have noted that children’s speech tends to not only be strikingly more egocentric than that of adults (Flavell et al., 1968), but children’s dialogues also tend to take the form of “serial monologues,” in which children take turns at speaking, but often fail to produce speech that follows the thread of the previous speaker (Piaget, 1926/1955).

Although many high-level decisions speakers make may be governed by common ground, our review has suggested that many of their lower-level decisions are surprisingly insensitive to common ground. Speakers do include information in their speech that a listener could not readily infer, such as mentioning an atypical instrument, but this appears to be a by-product of the way concepts are accessed during language production (Brown & Dell, 1987). Such adaptations may be made in regard to the needs of the particular listener (Lockridge & Brennan, 2002), but these particularized adaptations appear to be other-prompted and not self-prompted. Perspective adjustment could accommodate this latter finding by assuming that speakers who failed to mention atypical instruments
received negative feedback from addressees, which made them monitor their speech more closely. Furthermore, a multiple-systems approach seems necessary to explain the findings that, if individuals are placed under cognitive load, they fall back on a default form of speaking that is not tailored to the addressee (Horton & Keysar, 1996; Rossnagel, 2000).

Although speakers do appear to monitor aspects of their speech for audience design, they are not very proficient at detecting ambiguity in their own speech, even when they are not under load (Ferreira et al., 2005; Keysar & Henly, 2002). This distinction between use of common ground in the conceptualization versus formulation of an utterance plan poses problems for both full and partial constraint models, and supports multiple-systems approaches. Speakers do not use optional words or constituent ordering to avoid garden paths. In spoken dialogue, speakers do appear to adopt a prosodic marking that correlates with the syntactic structure, and these cues do appear to help listeners (Kraljic & Brennan, 2005; Schafer et al., 2000; Snedeker & Trueswell, 2003). But given that speakers use the cues independently of the listener’s needs (Kraljic & Brennan, 2005), the cues are not part of audience design. In addition, the articulatory reduction of repeated words in dialogue does not depend upon the listener’s perceptual needs, but appears to be the result of repetition priming (Bard et al., 2000; Bard & Aylett, 2005).

The finding that speakers use the definite and indefinite article to mark accessibility for the listener (Bard & Aylett, 2005; Hupet & Chantraine, 1992; Lockridge & Brennan, 2002) reflects a particularized adaptation. To explain this, perspective adjustment would have to assume that the selection of a definite versus indefinite form is first done egocentrically (i.e., on the basis of the given/new status of the referent for the speaker) and then adjusted during a later-monitoring stage. The dual process model provides a more parsimonious explanation by assuming that the marking of definiteness is a controlled process that figures into the initial design of an utterance. It is not clear how these two possibilities could be distinguished, since both theories would predict an “egocentric” result under cognitive load.

For language comprehension, the results clearly reject the full constraint model, since effects of private knowledge on comprehension have now been widely replicated in many studies by various laboratories, using a variety of tasks. The ordinary memory view also cannot explain the majority of the findings, because comprehenders appear to honor the distinction between private and common information. Instead, the bulk of evidence that we reviewed on language comprehension could be construed as supporting either perspective adjustment or partial constraint. Notably absent from the literature to date have been discussions of a dual process model for comprehension, although it could also explain the body of findings. More fine-grained, process-level investigations are needed to distinguish between these three competing models.

The abbreviation of description is another case in which particularized adjustments appear to be made; but here, too, multiple explanations exist. Abbreviation depends upon the presence of feedback (Hupet & Chantraine, 1992, 1994; Krauss & Weinheimer,
1966), but little is known regarding just how this feedback makes a difference. Feedback might enable the orderly accumulation of common ground (Clark & Brennan, 1991), and the shortening may be due to the speaker’s self-prompted use of this common ground. Alternatively, the speaker’s reduction in description length might be other-prompted, falling out naturally from the speaker’s sensitivity to evidence of listener understanding: speakers may terminate their descriptions as soon as they have evidence that the listener understands, which would come earlier and earlier at each reference turn.

Evidence regarding the use of conversational precedents in production and comprehension does not unequivocally support partner-independence or partner-specificity. Some of the partner-specific effects might be interpreted as other-prompted adaptations, such as speakers’ reversion to basic-level terms in Brennan and Clark (1996). Others, such as listeners’ spontaneous use of common ground when precedents are broken (Metzing & Brennan, 2003), would seem to be self-generated, posing problems for perspective adjustment. In contrast, the robust finding of no effect of common ground when precedents are maintained poses problems for both full and partial constraint models. In sum, no current model is able to explain the totality of the findings on conversational precedents.

Our verdict is that the multiple-systems approaches fare best in explaining the full set of findings reviewed in this chapter. The balance of findings would appear to tip slightly in favor of the dual process model over perspective adjustment, only because of a single finding: that speakers use the definite versus indefinite article according to the accessibility of referents for the listener, a finding that is explained more parsimoniously by the dual process model. Otherwise, the two models can account for the data equally well. Partial constraint can explain some of the findings – especially those pertaining to language comprehension – but encounters difficulties explaining how manipulations of cognitive load might affect audience design processes in language production, while leaving other aspects of production relatively intact. Finally, the full constraint and ordinary memory models can be rejected; the former based on the observation of effects of private knowledge on all aspects of production and comprehension, and the latter based on the observation that decision processes clearly honor the distinction between private and common information.

6.2. Interactive Alignment

The majority of the work reviewed here appears to undermine the proposal that speakers and listeners solve coordination problems by processing language against their common ground. There may still be important roles for common ground in very high levels of planning during language production, in the detection and correction of errors in both production and comprehension, and in knowing how to interpret feedback from one’s interlocutor. However, audience design does not appear to play a role in many of the routine processes involved in speaking and understanding. This might seem theoretically problematic, since egocentric processing would seem to provide a poor foundation for
cooperative communication. However, audience design is not the only possible explanation for why conversation succeeds. We close this chapter by considering an alternative proposal from Pickering and Garrod (2004).

According to the interactive alignment approach (Pickering & Garrod, 2004), listeners tend to understand speakers because mechanisms exist that promote the generation of cognitive representations that are shared, or in “alignment.” This alignment is not serendipitous, nor is it explicitly negotiated by interlocutors, but emerges as a by-product of the many individual acts of coordination that take place during conversation. Importantly, these shared representations are different from common ground in that they do not involve an explicit model of the other, nor are they necessarily “known to be shared.” The alignment process is facilitated by automatic priming processes, both within the individual language user as well as across individual language users. For instance, a speaker who produces a particular syntactic structure not only becomes more likely to repeat it in the future, but the comprehender who recovers that structure during the course of interpretation will also be more likely to produce it in the future. In this way, interlocutors’ representations converge over the course of a dialogue. Alignment is made possible by the assumption of representational parity: that the production and comprehension of language draw upon the same cognitive representations, a notion similar to the “parallelism” proposed by Dell and Brown (1991). As in perspective adjustment, common ground has a limited role in the coordination of shared understanding by enabling a process of “interactive repair” that can correct misunderstandings caused by insufficient alignment.

The findings reported in this review are largely consistent with the claims of the interactive alignment model. The model also enjoys a broad range of empirical support from studies showing spontaneous alignment between interlocutors in the absence of explicit negotiation, at many levels of processing, including speech rate (Webb, 1969), syntax (Bock, 1986; Branigan, Pickering, & Cleland, 2000; Levelt & Kelter, 1982), lexical choice (Garrod & Anderson, 1987), and conceptualization (Barr, 2003; Garrod & Anderson, 1987; Garrod & Doherty, 1994; Markman & Makin, 1998). Thus, interactive alignment offers a viable alternative for explaining coordination in language use. However, it is currently lacking a description of how this convergence of representations takes place. Although many aspects of alignment may be supported by automatic mechanisms such as priming, it may be necessary to invoke other mechanisms to explain certain aspects of language use (Brown-Schmidt & Tanenhaus, 2004; Kaschak & Glenberg, 2004; Krauss & Pardo, 2004; Markman, Kim, Larkey, Narvaez, & Stilwell, 2004; Shintel & Nusbaum, 2004).

6.3. Concluding Statement

The main lesson of our review is simple but important: when communication succeeds in the face of ambiguity, it does not necessarily entail audience design. Even if a speaker says something that seems designed for a listener, it does not mean that the speaker
actually designed it for that listener. Likewise, just because a listener understood an utterance as intended does not mean that she/he consulted her/his common ground with the speaker. Interlocutors may understand one another very well when neither of them uses common ground. After all, the cognitive representations that speakers and listeners tend to be working with are representations that are likely to be shared (Barr & Keysar, 2005; Pickering & Garrod, 2004).

Developing a theory of how people adapt their language behavior to meet the informational needs of other individuals is one of the central goals of research on language use. To meet this goal, researchers must surmount the challenges posed by the complexity of language use, a complexity that derives from its simultaneous cultural, interactional, and cognitive nature. Although the research of the past few decades has made great strides, understanding the joint activity of conversation is a task that will require the joint activities of researchers for many decades to come.

REFERENCES


CHAPTER 23. PERSPECTIVE TAKING AND THE COORDINATION OF MEANING


CHAPTER 23. PERSPECTIVE TAKING AND THE COORDINATION OF MEANING


