Title
Toward a Comprehensive Theory of Audience Design

Permalink
https://escholarship.org/uc/item/3d63r0qb

Author
Gann, Timothy Matthew

Publication Date
2012

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Toward a Comprehensive Theory of Audience Design

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Psychology

by

Timothy Matthew Gann

December 2012

Dissertation Committee:

    Dr. Christine Chiarello, Chairperson
    Dr. Curt Burgess
    Dr. Steven Clark

The Dissertation of Timothy Matthew Gann is approved:

_____

_____

_____
Committee Chairperson

University of California, Riverside

Acknowledgments

I would like to take this opportunity to thank the small army of undergraduate research assistants who helped out by putting in many hours of running participants and coding data: Chelsey Anderson, Justin Do, Huy Hoang, Taiki Kondo, Susanna Luu, Adrian Molina, Ashley Rodolf, Yoshiko Rosemond, Sarah Warner, and Jessica Rose Zendejas. A special thanks to Amy Dell for helping produce the stimuli used in the final experiment. Also, thanks go out to Mandana Seyfeddinipur for supervising some data collection.

The text of this dissertation, in part or in full (primarily Chapter III), is a reprint of the material as is appears in Language and Cognitive Processes (available online as of March 2012), published by Taylor & Francis. The co-author, Dale Barr, listed in that publication directed and supervised the research which forms the basis for this dissertation. Without Dale's help, none of this work would have been possible.

On a personal note, I don't think I would have survived graduate school if it wasn't for my friend, Russell Pierce. I also appreciate the support of Kauyumari Sanchez, who helped more than she probably thinks. A special thanks go out to Faye Harmer for her patience and support. I would not have attended UCR at all, and I certainly wouldn't have finished this dissertation, if it wasn't for my advisor, Christine Chiarello. Final thanks go out to the two other members of my dissertation committee: Curt Burgess and Steven Clark.

Dedicated to the memories of two women I wish could have seen this part of my education to its conclusion: my mother, Gay Nickole Gann, and my first graduate school advisor, Stella Royce Arambel.

ABSTRACT OF THE DISSERTATION

Toward a Comprehensive Theory of Audience Design

by

Timothy Matthew Gann

Doctor of Philosophy, Graduate Program in Psychology
University of California, Riverside, December 2012
Dr. Christine Chiarello, Chairperson

The purpose of this dissertation is to extend our knowledge of how speakers plan what they are saying to meet the needs of their audience. Three experiments were conducted to help connect our understanding of audience design with general cognitive processes of attention, memory, and learning. The first experiment tested how quickly speakers adapt to new common ground, and whether this adaptation is related to their eye gaze. It found that speakers often used speech repairs to integrate new common ground information indicating that they had difficulty adjusting their perspective. The second experiment focused on the question of whether audience design can be seen as a form of expert performance, and under what circumstances speakers are more or less likely to rely on this expertise. Speakers were observed relying heavily on memory routines established during a training phase, resulting in frequent misspecification of referents. In addition, self and other prompted speaker adaptation were considered in light of the potential for feedback by the addressee, and evidence was found that utterance planning was influenced by the interactional affordances of the situation. The final experiment examined how speakers learn information about their partners' perspective over time, and how this learning is impacted by the availability of feedback and role

constancy. Successful speakers integrated their partner's privileged knowledge, mostly through direct feedback but to a lesser degree also from the experience of taking turns as the director and addressee. Taken together, these experiments demonstrate that successful audience design is best understood as a process that weaves together many strategies and mechanisms. These range from a strong reliance on established linguistic routines to strategies such as selectively attending to privileged information and the opportunistic use of feedback.

Contents

List of Figures

List of Tables

Chapter 1

General Introduction

Conversation is the foundational, and most frequent application of language (Clark & Wilkes-Gibbs, 1986). The ease and fluency that characterizes most everyday conversation obscures the fundamental complexity of the process of producing even the simplest verbal expressions. This complexity goes beyond the problems inherent in the transformation of thought into action within an individual. Conversation is fundamentally social and cannot be fully understood without understanding it as a problem of coordination between two or more interlocutors. So while fruitful research can be done on the production and comprehension of utterances in isolation, language can be best understood as a joint activity (Clark, 1996) that is greater than the sum of its parts.

Understanding how people adapt their language to meet the needs of their partners is an important area of study because of its centrality to the human experience, and indeed having the capacity to have an awareness that others have knowledge and beliefs different from one's own (frequently called *theory of mind*) is often cited as one of the fundamental socio-cognitive differences that sets human cognition apart

(Tomasello, 1999)–though recent work by Crockford, Wittig, Mundry, and Zuberbühler (2012) suggests that chimpanzees have at least some capacity for taking the differential knowledge of group members' awareness and attention to potential danger.

One of the fundamental applications of conversational interaction is to allow a speaker and addressee to establish joint attention to a referent. For example, if a couple are in a kitchen preparing dinner and one asks the other for a spoon the goal of the speaker is to draw the attention of their partner to a needed utensil (and, ideally, retrieve it). Successful resolution of the speech act doesn't come merely with the speakers' drawing of the listener's attention to the spoon, but also with their mutual knowledge of shared attention to the intended referent (Clark, 1996; Tomasello, 1999). In all but the most trivial cases, this is more challenging than it first appears. In order to establish this joint attention, the speaker must satisfactorily identify the intended referent, by disambiguating it from possible competitors both from their own perspective and that of their addressee (Clark & Marshall, 1981; Olson, 1970), and construct an utterance that will convey this information. Perhaps there are multiple spoons available to choose from; perhaps there is a particular intended spoon in a place familiar to the speaker but not to the listener. Thus tailoring speech to meet the needs of one's audience quickly becomes one of the most important aspects of referential communication. The processes by which a speaker crafts an utterance that is sensitive to the needs of their addressee is called audience design (Clark & Murphy, 1982) or recipient design (Sacks, Schegloff, & Jefferson, 1974).

1.1 Audience Design

Given its central importance to successful communication, explaining how audience design is accomplished is critical for our understanding of dialogue. An early account of audience design focused on the concept of *common ground*, which is a form of meta-cognitive knowledge about what information is shared between conversational partners, and what they mutually know to be shared between them (Clark & Marshall, 1981; Clark, 1996). According to this view, speakers take advantage of evidence derived from shared sources of information including physical co-presence, linguistic co-presence, and community membership. Physical co-presence offers the greatest evidence of potential referents that are shared, not just because of the ease of inference regarding the shared nature of the immediate environment, but also because of the ease with which people can confirm a partners' mutual knowledge by attending to their gaze (Tomasello, 1999). Linguistic co-presence is established over the course of a conversation, and allows us to assume that what has been said previously is shared. Finally, community membership refers to information that can be inferred to be in common ground based on mutual knowledge of membership in communities that involve some particular sets of knowledge or experiences. Clark (1996) suggests that this is can be accomplished because interlocutors maintain mental models of what their addressees know.

Initial formulations of this view held that considerations regarding common ground influence every stage of linguistic processing (Clark, Schreuder, & Buttrick, 1983). In principle, it is possible for speakers to always deliberately design their utterances to be well formed in regards to the interlocutors' common ground, but such a strategy seems to be at odds with the speed of conversation and the types of errors and repairs that are typically seen. Given evidence that audience design processes can be

potentially costly, and subject to interference (Horton & Keysar, 1996; Roßnagel, 2000), this account of it as an obligatory part of production seems unsatisfactory.

One class of models regarding language production focuses on accessibility as the primary consideration in regards to whether information is included in an utterance. This suggests that the form and content of utterances is not driven purely by inference and deliberative processes, but is sensitive to the salience of particular concepts in the mind of the speaker (Brown & Dell, 1987; Wardlow Lane, Groisman, & Ferreira, 2006; Wardlow Lane & Ferreira, 2008). This is supported by a twist on a classic referential communication game as demonstrated by Wardlow Lane et al. (2006). A classic arrangement for a referential communication game is to present a speaker with two objects of the same basic category that contrast with each other along some dimension, and for the speaker to describe the target object such that their addressee can pick it out. If for example, a speaker is presented with a large toy giraffe and a small toy giraffe and told to describe the first giraffe, they will usually describe it as "the large giraffe." However, if you make it clear to the speaker that their addressee doesn't know about the smaller giraffe (through manipulation of physical co-presence), they will usually just say "the giraffe" (Clark et al., 1983; Horton & Keysar, 1996) following the Maxim of Quantity, which suggests that generally people will not supply more information than is required Grice (1975). In this case, if the addressee is not aware of a second object, speakers will typically not include information that draws a contrast with it. Wardlow Lane et al. (2006) conducted an experiment in which speakers were asked to describe an object to an addressee such that the addressee could both pick out that object and also so that they couldn't guess the identity of an additional object only in view of the speaker. This should not be a problem given that under normal circumstances speakers don't have a problem not including the additional information. However, because of the additional

instruction raising the salience of the concealed object speakers found it more difficult to not include the information–even though that was part of the task.

Similarly, Horton and Gerrig (2002) describe a comparable approach in which speakers are said to rely on ordinary memory processes in order to determine the content of their utterances. This hearkens back to the principle of encoding specificity (Tulving & Thomson, 1973), in which memory retrieval is linked to the environment and conditions under which it was initially encoded. Horton and Gerrig (2002, 2005a, 2005b) suggest this as an explanation for the existence of partner-specific influences on production, as partners act as cues for the retrieval of information associated with previous interactions without the need for an appeal to an exhaustive personal diary as described by Clark (1996). However, results from Brennan and Clark (1996) suggest that speakers do not immediately adapt the form of their utterances for a new conversational partner, but rather this adaption happens over a few trials and is not necessarily immediate. This doesn't suggest that partner-specific cues aren't an important aid in memory retrieval for previous referential descriptions, but that whatever benefit is imparted may be obviated by substantial past experience making a response particularly salient. These two effects can be consistent with speakers requiring a few trials to adjust what is salient to a novel conversational partner.

Both the accessibility and memory accounts offer a way in which apparent audience design can be accomplished without the need to engage an algorithmic audience design approach. Rather, information that is beneficial for a partners' understanding (or detrimental depending on the circumstances) can be incorporated automatically. This view coincides with the observation that information that is physically co-present, linguistically co-present, or involved with community membership is also very likely to be already active and accessible in memory, and likely to be incorporated into an

5

utterance without the need for audience design to be achieved through more involved consideration of common ground.

Other accounts of audience design are the Monitoring and Adjustment model (Horton & Keysar, 1996), the Dual Process model (Bard & Aylett, 2005), and the Anchoring and Adjustment model (Epley, Keysar, Van Boven, & Gilovich, 2004), which assume that production and audience design can be explained using both automatic and controlled components. The automatic process is thought to be egocentric, in which an initial formulation of what is to be said is generated with regard to information that is available to the listener, and this initial formulation is adjusted through a controlled process such as monitoring. To demonstrate this, Horton and Keysar (1996) used an experimental setup similar to the prototypical referential communication experiment described above. However, they manipulated cognitive load by putting the speaker under time pressure to begin their utterance. While under time pressure, speakers no longer followed the maxim of quantity and described the target without an adjective, but rather they tended to give an egocentric response that described the object from their own frame of reference. Experiments that find evidence for egocentric behavior tend to not give speakers much time to preview stimuli before production needs to begin (e.g. Horton & Keysar, 1996), or put speakers under cognitive load (Roßnagel, 2000, 2004). Experiments tend to show more early effects of common ground if they incorporate longer previewing times and more enduring stimuli to afford the chance to redirect attention away from privileged information (a competitor seen only by the speaker) before utterance planning. Indeed, most referential communication experiments that incorporate eye-tracking show less orienting to privileged information. Roßnagel (2000) also found a similar link between cognitive load and the production of more egocentric utterances.

These different sets of models do not necessarily have to compete however, since they address different levels of processing. The accessibility account offers a good framework for what information is automatically supplied to the formulation process, while the monitoring and adjustment account offers an account of how attention and monitoring is brought to bear on that initial formulation and refines it. One problem with the monitoring and adjustment account is that it could not fully explain the presence of an apparent early influence of common ground on production. The accessibility account supplies a solution by pointing out that highly accessible information, especially with some prior history or preparation, is as accessible as it is because it is very likely to be in common ground (due to co-presence or other shared knowledge). This should also make monitoring and adjustment more compatible with constraint-based models that suggest that common ground should have an early influence on production, and exert that influence at every stage of processing (Hanna, Tanenhaus, & Trueswell, 2003).

1.2 Fitting audience design into a general model of production

To further consider how to develop a comprehensive model of audience design, it is important to incorporate the models outlined above into a general model of language production, and a model of conversational interaction. The production of referring expressions begins with a process of conceptual preparation in which speakers generate a pre-verbal message that forms a template for what the speaker intends to say. As discussed above, this message can either be generated through reasoning and deliberation, or it can be built quickly using concepts and syntactic structures that are highly accessible due to salience, recent history, or common ground. One mechanism by which

syntactic structures might be cued in a similar fashion is by viewing them as constructions (Tomasello, 2003; Goldberg, 2006). Message formulation is an incremental process that selects concepts from the speaker's situation model that are consistent with their communicative intention, arranges them in a sensible linear order, and activates the corresponding lexical concepts. This message then gets passed on to later formulation and articulatory processes (Levelt, 1989; Levelt, Roelofs, & Meyer, 1999; Guhe, 2007). These later processes, starting with lexical selection, tend to be increasingly subject to controlled processes (Bard & Aylett, 2005).

During the formulation and articulation process, speakers are able to monitor their own speech. Levelt (1989) and others (Levelt et al., 1999; Schegloff, Jefferson, & Sacks, 1977) emphasize that as speakers produce an utterance, they are also engaged in comprehending their own speech which allows for errors in conceptual formula and lexical selection to be corrected. This monitoring is part of why many errors, depending on their type, are caught and corrected by the speaker before any feedback from an addressee, as speakers have two opportunities to make repairs to speech prior to that point: both as they are engaged in an incremental production process, and also in the space between conversational turns (Schegloff et al., 1977). Relevant to a later discussion, these first and second position repairs can be referred to as self-initiated rather than other-initiated repairs, which are driven by the addressee. The possibility of mid-utterance repairs, and speech-error data, suggest that production as a whole is an incremental process (Levelt, 1989), and that speakers begin producing their utterances before they have verified whether their message is fully adequate given the informational needs of the audience (Pechmann, 1989). This was further demonstrated in the experiments by Horton and Keysar (1996) and Roßnagel (2000) in which speakers were less able to monitor their utterances.

To fully understand how audience design fits into language production, it is also important to consider how speakers structure information so that it may be deployed in speech. The situation model (Zwaan & Radvansky, 1998) is a multidimensional conceptual structure representing the goals, objects, and actors active in the mind of the speaker. It also includes the relevant dimensions of, and relationships among, those components (see available knowledge and local context in Guhe, 2007). Thus information that has been built into the current situation model of a speaker will supply the building blocks that conceptual preparation uses to construct a message. The form of the resulting utterance will be dependent on the structure of the situation model and the inter-relations of the constituent concepts. Salience effects such as those reported by Wardlow Lane et al. (2006) could be explained by task demands to include information in a situation model that would otherwise tend to be excluded, raising the chances of retrieval despite the need to suppress that information to complete the task.

The situation model is also built incrementally as speakers make inferences and encounter new information from their environment. Thus situation model building depends on how speakers are allocating their attention. As objects are attended to, they are recognized and categorized simultaneously (Logan, 2002) and the corresponding concept is incorporated into the situation model. As inferences are made, the resulting concepts are also incorporated in the situation model. In a conversational context, speakers are encountering a steady stream of new information in the form of feedback from their audience that can be also incorporated into their situation model. As this happens, previously incorporated concepts can be re-categorized in light of the new information. Changes in the structure of concepts that make up the situation model should also drive changes in how speakers are directing their attention, which leads to further changes to the situation model.

This dynamic also suggests that conversation can act as an engine driving conceptual change for both partners in a conversation. As each partner informs the other about what they know, they will frequently encounter situations in which they do not have a category that satisfies both of their needs. In this case, they will collaboratively construct a goal-oriented ad hoc category (Barsalou, 1983, 1991), or *socially emergent category* that corresponds to their joint perspective (Barr & Kronmüller, 2007). For example, when two people discuss where they wish to go out to eat they may form a category of acceptable restaurants on the spot that gets developed over the course of the conversation that is neither pre-existing in memory nor particularly enduring. This socially emergent category is then incorporated into their current situation models. The degree to which it persists and moves from being ad hoc to permanent depends on continued use and consolidation.

This has three important consequences. First, as conversation unfolds partners should begin to adopt categorizations of referents that are consistent with the needs of the conversation, and this should drive changes in the way that speakers attend to potential referents. Such categories should also be more likely to be activated in the presence of the original conversational partner due to partner specific priming and encoding specificity (Horton & Gerrig, 2002; Tulving & Thomson, 1973). Second, the resulting shift in attention should allow speakers to incorporate information about their partner's perspective into their situation model. Finally, this information is then made highly salient for the next turn in the discourse which should lead to a state of conceptual alignment in which both the speaker and listener have closely overlapping situation models (Barr & Kronmüller, 2007; Pickering & Garrod, 2004, 2006; Garrod & Pickering, 2004). This overlap subsequently allows interlocutors to allocate their attention in similar ways. The new categories should begin to converge as they begin to attend to

the same information and stop attending to discourse-irrelevant information, facilitating understanding and further easing the establishment of joint attention. As speakers re-categorize the information in their situation model, the resulting changes in conceptual structure should lead to alterations in how they refer to objects. Over time, speakers will restrict their referential domain to shared information, so that when conceptual preparation of an utterance begins, the information accessed will be consistent with common ground.

For example, a couple discussing what restaurant they should have dinner may collaboratively construct an ad hoc category corresponding to a set of acceptable dining establishments. If one partner rules out a particular set of ethnic restaurants, the other partner will have to adjust their set to match and subsequent consideration of possible referents that don't match this rule will be reduced. As this category becomes more firmly established, it becomes the basis for future negotiations. When one partner brings up dinner, the other is likely to have activated their representation of what places are mutually appreciated and will not consider places known to be outside that set.

This framework fits fairly well with the Interactive Alignment model of Pickering and Garrod (2004), which describes two ways that common ground information might be used in a conversation. The first is through the use of full common ground which is much like the process of the audience design hypothesis mentioned previously. The second is through the use and creation of implicit common ground. According to the Interactive Alignment model, implicit common ground is generated in dialogue through priming. As interlocutors converse, they are mutually priming each other in such a way that they are foregrounding the same information in each of their respective sit-uation models. This leads to a state of conceptual alignment as described previously. When interlocutors' situation models are aligned, speakers can then employ a strategy

of speaking egocentrically because there is a high probability of that their perspective overlaps with that of their partner. The processes described earlier are very similar to the development of implicit common ground. Both result in the speaker adopting a perspective that is similar to the listener's. Each also predicts that apparent effects of perspective should increase over the course of a conversation. The early part of a conversation may be characterized with relatively slower audience-design like processing until implicit common ground is established; later processing should be faster due to a reliance on overlapping attention and conceptual representations. However, the building of implicit common ground most likely involves any process that updates the situation model in such a way that its contents better reflect the content of the addressee's situation model. This may not be restricted to priming, but may also include implicit and explicit learning, and manipulation of the salience of the contents of the situation model through the manipulation of attention.

Conceptual overlap is significant because it allows interlocutors to minimize the effort involved in dialogue, which is one of the chief advantages of a cooperative interaction (Clark & Wilkes-Gibbs, 1986; Beun & Cremers, 1998). It has also been shown that speakers can rely on having significantly overlapping knowledge with their partner as a strategy to allow egocentric processing to proceed without causing miscommunications (Wu & Keysar, 2007b). This overlap first develops locally, such as in the current conversation (Garrod & Anderson, 1987), or between close associates (Fussell & Krauss, 1992). However, over successive conversations this overlap can spread to a community (Barr, 2004; Garrod & Doherty, 1994). The fact that we can communicate easily with people whom we have never met before is evidence for a widespread conceptual overlap between people in a language community. Speakers of the same language with similar backgrounds will typically name objects in the same way (Markman & Makin, 1998),

which also indicates that there is a considerable amount of implicit common ground for people to rely on.

1.3   Routinization and Feedback

A fully realized model of audience design needs to incorporate the strengths of the memory and monitoring and adjustment accounts discussed previously, such that they complement each other. Hearkening back to the memory view, it is useful to look at audience design as a form of expert performance in which speakers draw upon a large database of stored examples which they can consult and modify to suit their current needs (c.f. Simon & Chase, 1973; Ericsson, 2006). The Instance Theory of Automaticity (ITA; Logan, 1988) is a process model of learning that offers one account of skilled performance and automaticity (Proctor & Vu, 2006). In this model, decisions are instantiated as a race between competing alternatives, one based primarily on memory retrieval and the other on a more deliberative algorithmic process, with the swiftest solution being adopted. In terms of language production, as memory routines are developed over the course of a conversation, they are more likely to win this race because they are automatized, and thus made more swiftly available, through repeated use. This view fits in well with the Interactive Alignment model which also suggests that participants in dialogue build up a series of routines based on what they have said and heard previously (though chiefly through priming), that they can then deploy later in order to reduce effort (Pickering & Garrod, 2004, 2005). Addressee feedback is important in the process of building routines, since it is in the context of the dialogue that the success or failure of a particular routine can be judged.

While early treatments of audience design focused on the role of the speaker

in audience design, context and the addressee can play an important role in shaping the information present in an utterance (Clark & Wilkes-Gibbs, 1986; Schober & Clark, 1989). Feedback from the addressee does not have to occur during their turn in the conversation in order to have an impact, as speakers are able to continuously monitor the comprehension of the addressee through their behavior (Clark & Krych, 2004). This contributes to a slightly more fluid situation in which feedback and repair can occur earlier than previously supposed. It is possible to have an other-prompted adaptation earlier than the recipients turn since, for example, a speaker can potentially monitor their addressee for signs of comprehension while they are simultaneously engaged in production. However, the main point about self-prompted repairs tending to occur earlier than other-prompted repairs still holds. Feedback is important for allowing routines to be further sculpted so that speakers can adapt to the needs of the current interaction. For example, in Brown and Dell (1987) speakers tended to mention atypical instruments used in a narrative, but this behavior seemed to be driven by the speakers' perspective since the additional information was not relevant to the addressee. However, in that experiment the addressee was a trained confederate who did not necessarily respond the way a naïve listener would. In a more recent study (Lockridge & Brennan, 2002), it was found that, when a confederate is not used, speakers are more likely to adapt themselves to the addressee. This suggests that feedback is a critical element in the development of utterances.

Another reason for the importance of feedback is that it may allow speakers to engage in a strategy that reduces the cognitive burden of production and share it with the listener (Deutsch & Pechmann, 1982). Speakers may be capable in principle of engaging in a deliberate, but slow, process of audience design but may choose not to when circumstances permit. As speakers produce an utterance they can monitor the

14

reaction of the addressee and repair as necessary. In one sense, the addressee is the best possible model of what he knows and offers the most sure test of what is appropriate. In this sense, speaking can be construed as an *epistemic action*, an action taken to divest some of the required effort onto the environment (Kirsh & Maglio, 1994). Also, as the amount of overlap in the perspectives of the interlocutors increases, this strategy should be increasingly successful even without feedback since the egocentric perspective of the speaker felicitously matches the perspective of their audience. So the amount of feedback required should decrease over the course of a conversation even with such an opportunistic strategy in use.

1.4 Attention and Categorization

When speakers engage in referential communication, they are generating labels that will pick an object out from a set of alternatives such that the addressee can direct their attention to it (Barr & Kronmüller, 2007). As described above, one way of doing this is through the creation of a socially emergent category. How do these socially emergent categories compete with more established taxonomic categories (see Barsalou, 1991), and how do these new categories serve to direct attention away from privileged information and towards shared information?

One framework that can help answer these questions is a descendant model of ITA, the Instance Theory of Memory and Attention (ITAM; Logan, 2002). ITAM seeks to explain attention and categorization as two sides of the same process; objects are selected at the same time that they are categorized. In this model, there are several parameters that effect the outcome and influence which categorizations are chosen and where attention is directed: task-based factors, and historical factors. Task-based factors

can include the speaker's goals, any bias towards a particular categorization in support of those goals, and feature salience. Historical factors include the speaker's past experience with a category and the features associated with it. In this framework, socially emergent (ad hoc) categories can be construed as putting much emphasis on task based factors: they emphasize the need to select a novel categorization in order to achieve a goal, and emphasize certain features (like shared information) over others (privileged information). More established categories tend to be selected for a particular object because of past history and perceptual similarity to a category exemplar. What this means is that early on in a conversation, taxonomic categories should be activated more often and attract more attention. But as time goes by and speakers learn to put more emphasis on shared information, the socially emergent categories should become more prevalent.

People do learn to attend to features and objects that are relevant for completing their task, and ignore those that are not. Rehder and Hoffman (2005) conducted an experiment in which participants performed a category learning task while having their eyes tracked. At the beginning of their experiment, participants would attend to each of the features in the display and consider them all before making a decision. By the end of the experiment, participants learned to attend only to those features that were necessary for the task of choosing which category was represented by the display. If one was relevant, they learned to attend to just that one. If two were relevant, they learned to attend to only to two, and so on. In a referential communication experiment involving perspective taking, as time passes and socially emergent categories become more dominant, attention should shift away from privileged competitors that are not shared from the listener's perspective. By the end of a conversation these privileged competitors should be attended to about as little as an object unrelated to the target. This shift should also happen in situations in which the speaker does not initially have

16

access to what the addressee knows. As evidence about addressee's knowledge, or lack of knowledge, becomes available, they should gradually begin to focus on the stimuli that are most consistent with the addressee's perspective.

The question then becomes, what factors influence the rate of learning socially emergent categories and the content of these ad hoc categories formed during conversation? Two possibilities are feedback, and the roles of the interlocutors: are they in a didactic situation in which the roles are fixed, or is it a fully interactive dialogue in which there is turn taking? Feedback can come in many forms. First, speakers can receive pragmatic feedback in the form of knowledge about the success or failure of the addressee in establishing joint attention to a target object. It is implicitly accepted that pragmatic feedback is necessary for learning to happen, with positive feedback leading to a reinforcement of the speaker's current conceptualization, and negative feedback encouraging re-conceptualization. Second, the addressee can give the speaker signals from which they can infer the addressee's confidence about an interpretation. Barr (2003) found that addressees will use speaker's confidence as a cue that the current referent is either a typical or atypical member of a category. In principle, speakers should also similarly use the addressee's confidence in their selection of a target as a cue to how well their message was received.

Finally, addressees can give direct verbal feedback to the speaker. Schober and Clark (1989) conducted an experiment in which a participant acts as a non-interactive observer to a conversation. This participant attempted to do the same task the actual addressee was doing, but without the ability to give feedback to the speaker. They found that the observers never aligned as closely as the addressee to the speaker due to their inability to give feedback and have the speaker adjust to or correct their interpretations. The lack of verbal feedback can also make the speakers more prone to give

more information as a hedge against possible misunderstandings. For example, Krauss and Weinheimer (1966) found that when addressees could not give feedback, speakers tended not to simplify their utterances, but kept them long in order to not risk giving too little information. While it is possible for alignment to happen without it, direct verbal feedback likely serves to speed up the process. Without it, speakers may not be able to easily settle on a shared perspective and fully acquire a socially emergent category that corresponds to the intersection of their knowledge with their partner's.

Role switching may serve a similar function. In a natural conversation, both participants are speakers and addressees. In this case, each are offering their own conceptualization of the referents they are bringing to each others attention and are able to interactively sculpt each others understanding. They are constantly giving each other positive evidence about what they know, and have the ability to give corrective feedback. In situations where one person is the primary speaker, there might be less incentive for them to align in a symmetrical manner; in these cases the addressee may be expected, in light of their role, to preferentially align to the speaker's perspective.

While attention allows for the selection of information to be included into the situation model and reflect ad hoc category formation, the opposite process also happens. As the concepts tied to particular referents are determined to be less salient, less an integral part of the situation model, attention to them should decline. Referring again to the study by Wardlow Lane et al. (2006) where increased salience made the leaking of information inevitable (due to increased attention to the object to be concealed), the opposite process is observed: privileged information that is not made salient through some other means is attended to less.

1.5   The Current Investigation

This study aims to develop a theory of audience design that synthesizes the strengths of the antecedent models within a more explicitly interactive framework, that can explain what types of information become available, and when, and how appropriate information is selected for. In short, it attempts to combine the strengths of the memory and accessibility approaches with the strengths of monitoring and adjustment, within the broader framework of more general theories of interactive dialogue such as that described by Pickering and Garrod (2004) and what is known about broader cognitive mechanisms such as memory, categorization, and attention.

The experiment in Chapter 2 seeks to demonstrate that not only do speakers learn to direct attention away from privileged information, this redirection of attention can be correlated with increasing appropriate utterances in regards to common ground. This would tie together the accessibility account with the predictions of the interactive alignment model regarding to the influence of common ground as a process that unfolds over time rather than being based on isolated inferences. In this experiment, after speakers have learned to direct attention away from privileged information, the context is changed so that what was previously privileged is now shared. The question is whether speakers can adapt to the new context and formulate descriptions that take the new common ground into account immediately, or whether they demonstrate a continued tendency to ignore what was previously privileged. Wardlow Lane et al. (2006) put speakers in the position of leaking information that was intended to remain private by making it more salient. This experiment seeks to put speakers in a position of failing to disclose information that should be shared by making it less salient. This would further confirm the accessibility account and offer an additional strike against the initial design

hypothesis.

The second experiment, in Chapter 3, provides a test of whether the incremental nature of audience design relies on the use of routines. Speakers are put into a situation in which, after a period of training, they will have generated routine descriptions for a series of referents. The idea is that the speaker will either tend to give less information (under-specification), or more information (over-specification) than a naïve listener would require to identify it because they will rely more on the pre-established description regardless of whether it is optimal. In the test phase of the experiment, speakers describe the same targets but without the previous context, and with either the same or a different partner. The partner manipulation is meant to indicate whether or not speakers will adjust descriptions to the needs of a new partner, and what manner of repair they will employ to do it. Whether or not the speaker had the opportunity to receive feedback was also manipulated in order to test whether speakers were attempting to use an opportunistic strategy by using the availability of feedback to address deficiencies in their productions.

Chapter 4 implements a novel referential communication experiment in which both interlocutors have a privileged competitor in addition to a shared competitor. All of the competitors vary along dimensions that are consistent for each participant, so that both the shared dimension and the privileged dimensions are the same for each participant across all trials. The goal is to attempt to observe the development of a socially emergent category by observing the rate at which speakers and listeners adapt to each others' behavior and whether speakers tend to adapt to the perspective of the listener or vice versa. To the degree the speaker adapts her speech, one key question is whether the category she forms constitutes a union of her and the addressee's perspective, or an intersection in which the category only includes the shared dimension. As in

Experiment 2, feedback is manipulated in order to see whether it influences the speakers' tendency to either adopt or ignore the perspective of their partner. Additionally, partner roles were manipulated in order to see if interlocutors can better learn to adjust to their partners perspective by observing their speech and whether this would have a differential effect on the strategies employed by the participants.

Chapter 2

Experiment I

2.1 Introduction

Wardlow Lane et al. (2006) demonstrated that when privileged information is made to be highly salient by the task, speakers will tend to incorporate that information into their utterances. This is in violation of both the norm of giving no more information than is needed, but also of the task which required that they keep that information secret. The task demand that a particular piece of privileged information be concealed ironically made that information more accessible, and thus more likely to be automatically incorporated into an utterance by the language production apparatus. In combination with the studies by Horton and Keysar (1996) and Roßnagel (2000), this strongly suggests that speakers may start with an egocentric description of an object and adjust that description to be in line with their partner's needs. In the first study by Wardlow Lane et al. (2006), speakers leak privileged information because attention, and thus accessibility in memory, is focused on the privileged competitor. In the latter studies, it is because speakers are put under sufficient cognitive load to interfere with monitoring.

An important question that arises then, is how speakers manage interference from privileged competitors. Is it enough that they don't place any additional significance on the competitor, or do speakers tend to employ a strategy for reducing the chance that privileged information will be leaked under normal circumstances? In a study similar to that of Horton and Keysar (1996), Gann (2007) tested the effect of time pressure on the use of common ground information. However, in Gann (2007) whether or not participants knew the position of privileged information was manipulated. In the critical conditions the participants knew where the privileged competitor would appear before they knew what the objects would be. In this case, they less frequently included unnecessary privileged information in their utterances. The ability for speakers to suppress privileged information in this circumstance suggests that the Wardlow Lane et al. (2006) study not only made the privileged competitor more salient, but prevented the speaker from attempting to reduce interference from the competitor by giving speakers a continuing reason to focus on the competitor. In the previous studies it is hard to assess the effect of attention because of the lack of accompanying online measures such as eye-tracking data.

Being able to divert attention away from a competitor may present an avenue by which speakers may demonstrate early effects of common ground under less taxing circumstances, in addition to increased memory accessibility of common ground information. A strategy of attempting to minimize attention to irrelevant common ground information would allow a speaker to demonstrate apparent earlier effects of ground in the message formulation process. This would not be because they are incorporating common ground information into their messages at the time of planning, per se, but because the speaker may make the irrelevant information less salient and thus less likely to be incorporated into future utterances by virtual of its availability. A related

question is whether or not speakers can quickly adapt this strategy when the basis for common ground shifts, further differentiating whether or not the effects are due to the consideration of common ground or due to earlier attentional selection.

To test these ideas I constructed a referential communication experiment in which speakers were told what positions in the stimulus array contain objects guaranteed to be shared with their partners, and which were never be shared. This allowed a speaker, over the course of the experiment, to adapt to what areas they need to attend to and which areas they can safely ignore in order to avoid information that might be relevant from their point of view, but not from their partner's. This should allow them to proactively avoid attention to privileged information as in Gann (2007). In order to test whether this adaption to common ground is immediate and driven by the addressee's needs, or whether it is determined by how the speaker has learned to allocate his attention, the common ground status of the ignored areas was changed mid-experiment. The prediction was that speakers would not fully adapt to the new common ground status of objects they had previously avoided and would demonstrate a continued lower rate in the use of previously privileged information. In order to control for whether it is their privileged status (thus common ground) driving the redirection of attention, or the instruction that nothing will be seen in the areas designated as privileged, a baseline condition was implemented such that some participants never saw competitors in the privileged positions despite being instructed that those positions were privileged.

I expect that the effect of ground would grow over time as speakers became accustomed to the task and learned to differentiate between shared and privileged information based on position. Similarly, attention to privileged information should decline over this time course. However, after switching the context such that the previously

24

privileged positions are no longer privileged, I predicted that speakers would initially continue to treat those objects as privileged, in contrast to the view in which common ground should have an immediate effect. To the degree that privileged information is used, it should be as part of a repair–an addition after an initial formulation. If participants fail to correct themselves, it could suggest confusion over the task.

2.2 Method

2.2.1 Participants

Forty-eight participants from the University of California, Riverside contributed data to this study. All were recruited from introductory psychology classes and participated as a part of their course requirements. There were 28 females and 20 males in the sample. Every participant was clandestinely assigned to the role of director (speaker) for the experiment, while a trained confederate always fulfilled the role of the matcher (addressee) for each session. The confederate was a male of traditional college age, and was treated as if he was a naïve participant. During the participant debriefing, participants were asked if they thought the confederate had a pre-existing relationship with the experimenter. No participants included in the data set reported suspecting deception.

2.2.2 Apparatus

An ISCAN ETL-400 remote eye-tracker, sampling at a rate of 60 Hz, was used to track the director's gaze throughout the critical trials of the experiment. To aid in maintaining calibration of the eye-tracking equipment, the director was seated in a chair with a headrest against which they were asked to lean their heads and attempt to minimize movement. The experimental stimuli were displayed for the director on a

19" LCD monitor with a 4:3 aspect ratio, and the matcher's monitor was a 17" LCD monitor with a 4:3 aspect ratio. Each participant's screen was only visible to themselves. The director was given a set of headphones through which instructions could be given without the knowledge of the matcher.

2.2.3   Materials

Stimuli were presented within the context of 18 virtual 3x3 grids, each of which was used for two trials of the experiment. Examples of the director's and addressees' views can be seen in Figure 2.1 and Figure 2.2, respectively. This setup is reminiscent of physical grids used in studies such as Keysar, Barr, Balin, and Brauner (2000) and Wu and Keysar (2007a). The cells of the grids were identified by number prior to the presentation of stimuli for a given trial. Each grid included two targets (for the two trials), and potentially a competitor for each target that contrasted with its target along the dimension of size (either being smaller or larger relative to the target), and three objects unrelated to the critical items. The objects used as items for the experiment were an assortment of common artifacts, animals, and plants (see Appendix A for a list). The stimuli were all color images, obtained from internet clip art, set on black backgrounds so that they would blend in with the black background of the display. Two grids (4 trials) were set side for use for practice trials. Sixteen grids were used for the test trials; 12 for the training phase (24 trials/targets), 4 for the test phase (8 trials/targets). Each stimulus picture had width and height of 120 pixels.
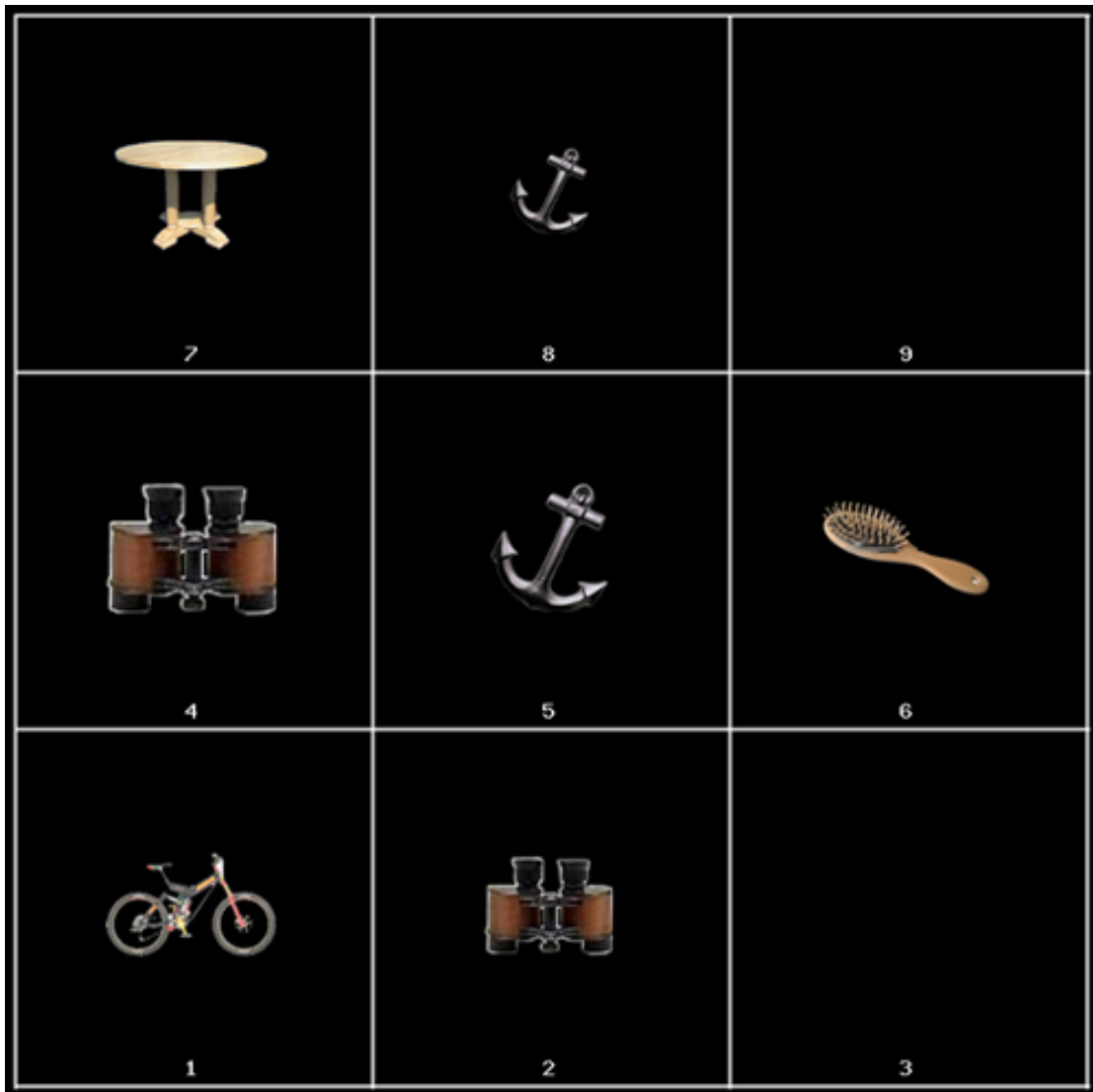
Figure 2.1: Directors are instructed that objects in the corners are privileged and are definitely not seen by the addressee. All other objects are potential targets, and there were two trials conducted per grid.
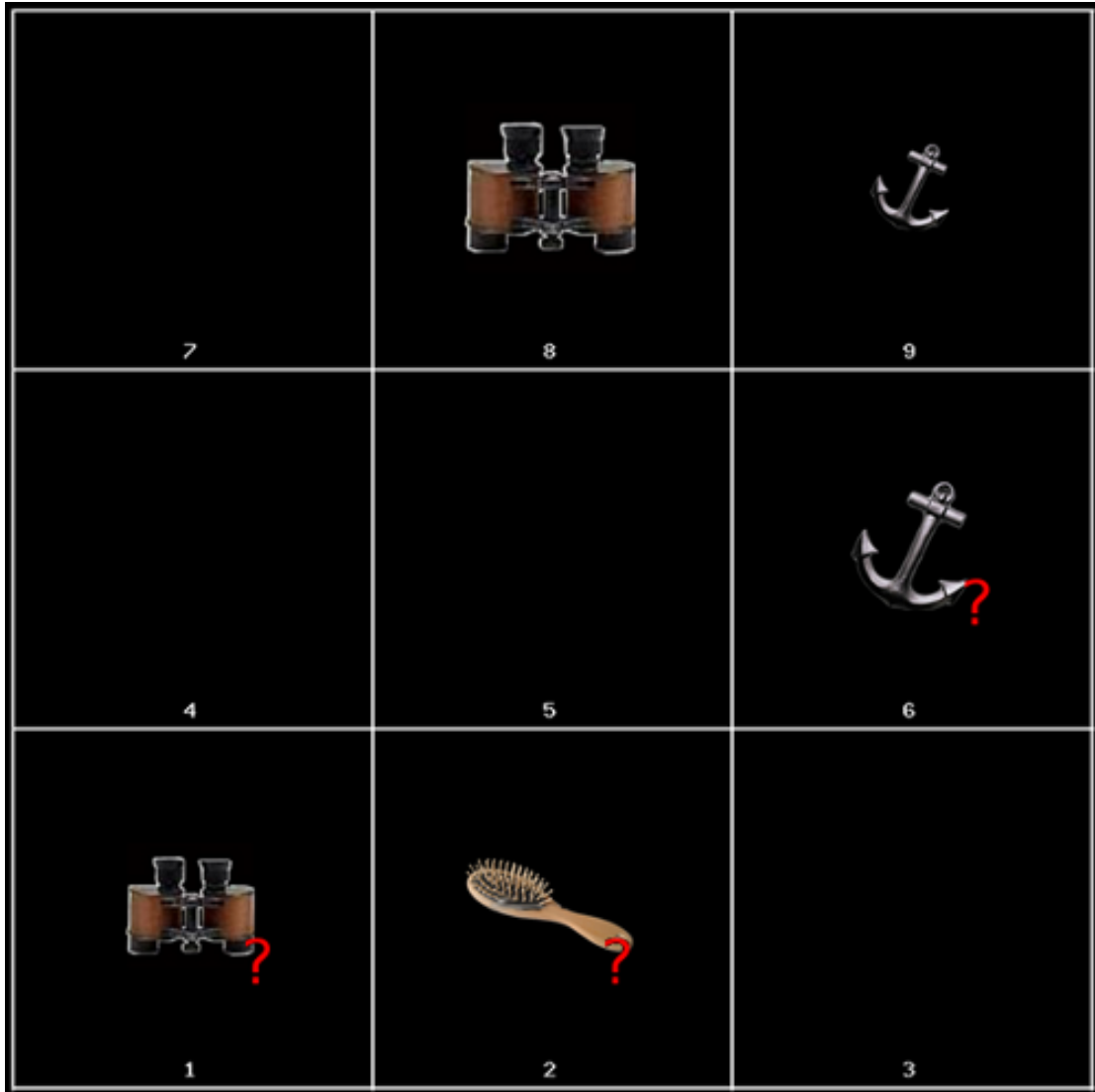
Figure 2.2: Addressees are not given any special instruction about the status of objects in the corner. However, objects marked with a question mark are ambiguous in regards to common ground and the addressee is instructed (as part of their cover) that these objects may, or may not, be seen by the director.

2.2.4 Procedure

The confederate was instructed to arrive slightly after the start of a scheduled session in order to reinforce the idea that the confederate was not affiliated with the experimenter. When both participants had arrived, the experimenter told the naïve participant to choose one of two cards at random to determine their role. Each card marked their role as the director, ensuring that they would be assigned to that role.

The participants were then instructed that they would be playing a cooperative matching game in which the director would be describing an object to the matcher, such that the matcher could pick it out on their display. The director was told that they would see a numbered grid mostly filled with objects, and that after a short preview (1500ms) they would be given the grid location of the target they had to describe through the headphones. The director was further instructed that the matcher is not able to see any object that appears in the corners of the grid and that those objects could be safely ignored. The director was also told that all of the other objects are guaranteed to be shared with the matcher. Finally, they were reminded not to use the grid number to identify the object for their matcher, as the object locations would be different for each participant.

The matcher was given a similar display. However, instead of being given the instruction that the corners are privileged, they had certain items marked with a question mark. This was meant to inform them that the directors may, or may not, know about that item. Any object not so marked was given as being shared with the director. Competitors for the matcher were made ambiguous in this fashion to set up a situation in which the director would be penalized for using privileged information. In trials in which the director sees a privileged set of small binoculars, the matcher would

see the target paired with a large set of binoculars marked with a question mark. If the director in this case uses the privileged information, the matcher will reason that the director does see the ambiguous set of binoculars, assume they are the target, and choose incorrectly on the basis of the unnecessary additional information. A confederate was used for the matcher's role after experiment piloting to ensure consistent performance due to the focus of the experiment on the director's speech production.

After the director gave their description, the matcher chose the object that they thought was being referred to. If they chose correctly, the targets were replaced with small flags. If they chose incorrectly, they were replaced with bombs. If they chose incorrectly on a trial with a privileged competitor, the director was reminded of the fact that the matcher cannot see objects in the corners. The participants were told that the goal of the game was to collect as many flags as possible.

After the instructions, the participants completed two practice grids (4 trials). After an additional 12 grids (24 trials) designated as the training phase, the experiment underwent a simulated crash that indicated that the matcher's screen was malfunctioning. The experimenter at this point made a statement to the participants about the apparent error, and instructed the participants to finish the rest of the trials in the interest of completing the data set with the matcher now standing next to the director and looking at the same screen. The participants were instructed that the fundamental task was the same, that the director merely needed to describe the target such that the matcher could pick it out, but now the matcher would select the object on the director's screen. Four additional grids (8 trials) were then conducted during this test phase.

2.2.5 Design

The design of the experiment included one between subjects factor, baseline vs treatment group. The treatment group had privileged objects included in the corners, whereas the baseline group did not. Thus the baseline group represented a condition in which there was no functional reason to ignore the corner squares. The test phase of the experiment always had trials that included "privileged" (in the corners) competitors for both the baseline and treatment groups. Ground constituted a within subjects factor, which in the treatment group was a comparison between trials that had a shared competitor and those that had a privileged competitor. In the baseline condition, this comparison was between trials that had a shared competitor and those which had no competitor. Trial order was also a within subjects factor, in order to allow for the analysis of the effect of time. The training and test phases were analyzed separately due to the change in the meaning of the ground effect between them. The dependent variables were accuracy (whether or not the addressee chose the correct target), modifier-use (whether a size adjective was used), modifier position (was the adjective positioned before or after the noun), speech onset time, and proportion of time spent fixating objects in the privileged corners of the display.

2.2.6 Analysis

The director's speech was recorded, transcribed, and coded for accuracy, speech onset, the use of a size modifier, and whether that modifier was placed before or after the first instance of the noun corresponding to the target for each trial. Coding was done by two undergraduate research assistants who were blind to condition, with the experimenter resolving any differences (while also being blind to the condition). The

eye-tracking data was coded as number of frames (sampled at 60 hz) that the director spent looking at each object prior to speech onset.

The analyses were conducted using linear mixed effects models with random effects included for subjects and items (Baayen, 2008). Models were fit within R using the lmer function within the package lme4 version 0.999375-39 (Bates & Maechler, 2010). Appropriate link functions were chosen depending on the distribution of the outcome variable of interest, as noted below. Random effects corresponding to the within-subject variables of ground and trial order were included in the subjects random effect portion of the models, and retained if they explained significant amounts of variance in the model as tested by a χ^2 model comparison using the anova function in R. A similar procedure was used to find the ideal random effects structure for the item random effect. Similarly, the significance of the fixed effects were assessed using the same χ^2 model comparisons approach (Barr, Levy, Scheepers, & Tily, 2013).

2.3   Results

Overall, speakers were successful at guiding their partners into selecting the correct referent, averaging 88.77% accuracy across all conditions. A total of 40 trials were removed from the data set out of 1536 (2.6%) due to protocol violations (non-trial related speech at the start), or errors in the experimental program. For a summary of the data on trial accuracy, consult Table 2.1 and Figure 2.3. In the critical test phase, after the simulated experiment crash, trial order was an important factor in determining accuracy ($\chi^2_{(1)} = 4.302, p = 0.038$). Crucially, ground also played an important role at the beginning of the phase ($\chi^2_{(1)} = 6.999, p = 0.008$); accuracy was lower for privileged trials (56.52%) than for shared trials (87%). This demonstrates that when the formerly

Table 2.1: Inferential statistics for accuracy in the test phase, first for Order as a mean-centered variable (Ground refers to the effect at the midpoint of the experiment), and second for Order (first trial coded as 0 so that Ground is interpreted as the effect at the start of the phase).

Factor	Initial Effects			Centered Effects		
	Estimate	$\chi^2_{(1)}$	p	Estimate	$\chi^2_{(1)}$	p
Ground (G)	1.969	6.999	$< 0.008^{**}$	0.600	1.538	0.215
Baseline (B)	-0.469	0.281	0.596	0.672	0.824	0.364
Order (O)	0.253	4.302	0.038^{*}	”	”	”
G:B	1.153	0.566	0.452	-1.629	3.072	$0.080^{\dagger}$
G:O	-0.391	3.254	$0.071^{\dagger}$	”	”	”
B:O	0.326	2.758	$0.097^{\dagger}$	”	”	”
G:B:O	-0.795	3.353	$0.067^{\dagger}$	”	”	”

$^{\dagger}p < 0.10$; $^{*}p < .05$; $^{**}p < .01$.

privileged items in the display are brought into common ground, speakers initially have difficulty in guiding the listener to the appropriate referent. This result is in line with the hypothesis that speakers would struggle to adapt to the new common ground after the training phase. An important factor to consider in regards to this performance, is that the listener is a confederate trained to respond in a uniform manner based on the speaker's input. So while accuracy is expected to be intimately tied to the form of the utterance produced by the speaker, this connection should be even greater due to the presence of the confederate.

While the use of modifiers in the test phase is key, it's important to note that the training data in the test condition conformed to expectations (refer to Table 2.2 and Figure 2.4). Over the course of the training there was a significant effect of ground ($\chi^2_{(1)} = 328.45, p < 0.001$), which existed from the first trials ($\chi^2_{(1)} = 19.604, p < 0.001$). There was also a significant interaction between ground and trial order ($\chi^2_{(1)} = 26.803, p < 0.001$) indicating that the ground effect grew in size as speakers became accustomed to the experiment and adapted to their listener's needs. By the end
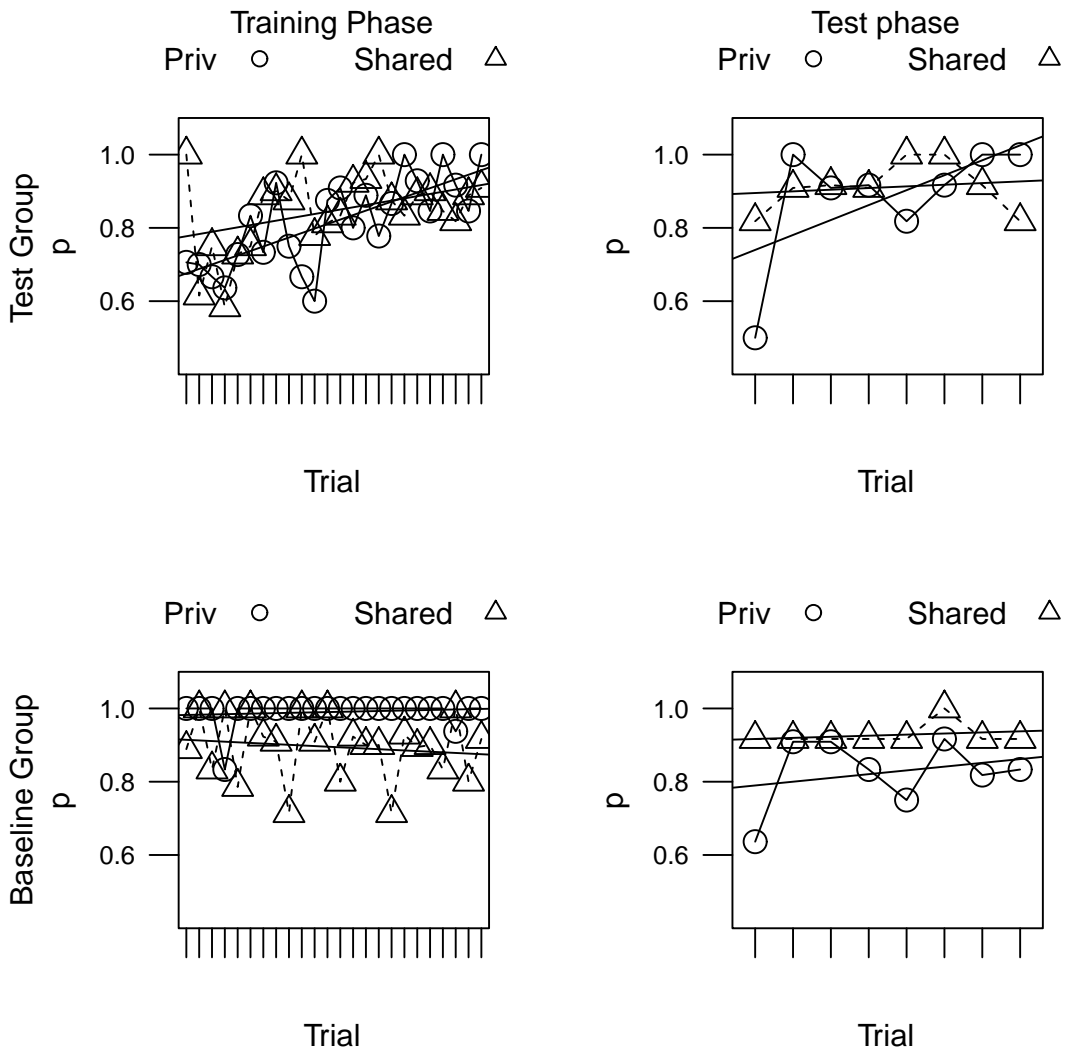
Figure 2.3: Accuracy plotted as a proportion of correct responses by the addressee.

Table 2.2: Inferential statistics for modifier use in the training phase, first for Order as a mean-centered variable (Ground refers to the effect at the midpoint of the experiment), and second for Order (first trial coded as 0 so that Ground is interpreted as the effect at the start of the phase).

Factor	Initial Effects			Centered Effects		
	Estimate	$\chi^2_{(1)}$	p	Estimate	$\chi^2_{(1)}$	p
Ground (G)	2.270	19.604	$< 0.001^{**}$	4.978	328.45	$< 0.001^{**}$
Order (O)	0.011	0.226	0.128	"	"	"
G:O	0.235	26.803	$< 0.001^{**}$	"	"	"

$^{\dagger}p < 0.10$; $^{*}p < .05$; $^{**}p < .01$.

of the training, modifier use was almost perfectly aligned with the ground condition, setting up a clear point of comparison with the initial trials of the test phase.

The critical prediction for the test phase was that the speakers, when confronted with a situation in which information they had been trained to discount was no longer privileged, would initially continue to treat the formerly privileged information in the same way. Thus, while in the test phase all trials were in actuality "shared" by virtue of the listener viewing the same screen as the speaker, the expectation was that speakers would continue to treat the corners as privileged. This prediction was born out: At the beginning of the test phase there was a significant difference due to ground ($\chi^2_{(1)} = 13.993, p < 0.001$), use of modifiers in the shared condition remained high (90.9%) while in the formerly privileged condition they remained lower (58.3%). Refer to Table 2.3 and Figure 2.6.

Although modifier use in the privileged condition did not follow the ideal of initially staying at the same low level seen in the training phase, the substantial 30% difference between the conditions at the start of the test phase still supports the prediction. This difference comes even closer to mirroring the performance from the end of the training phase if one considers the possibility that speakers adapted to a delay in re-

35

Figure 2.4: Proportion of trials in which the director incorporated size modifiers (irrespective of modifier position) into their utterances.

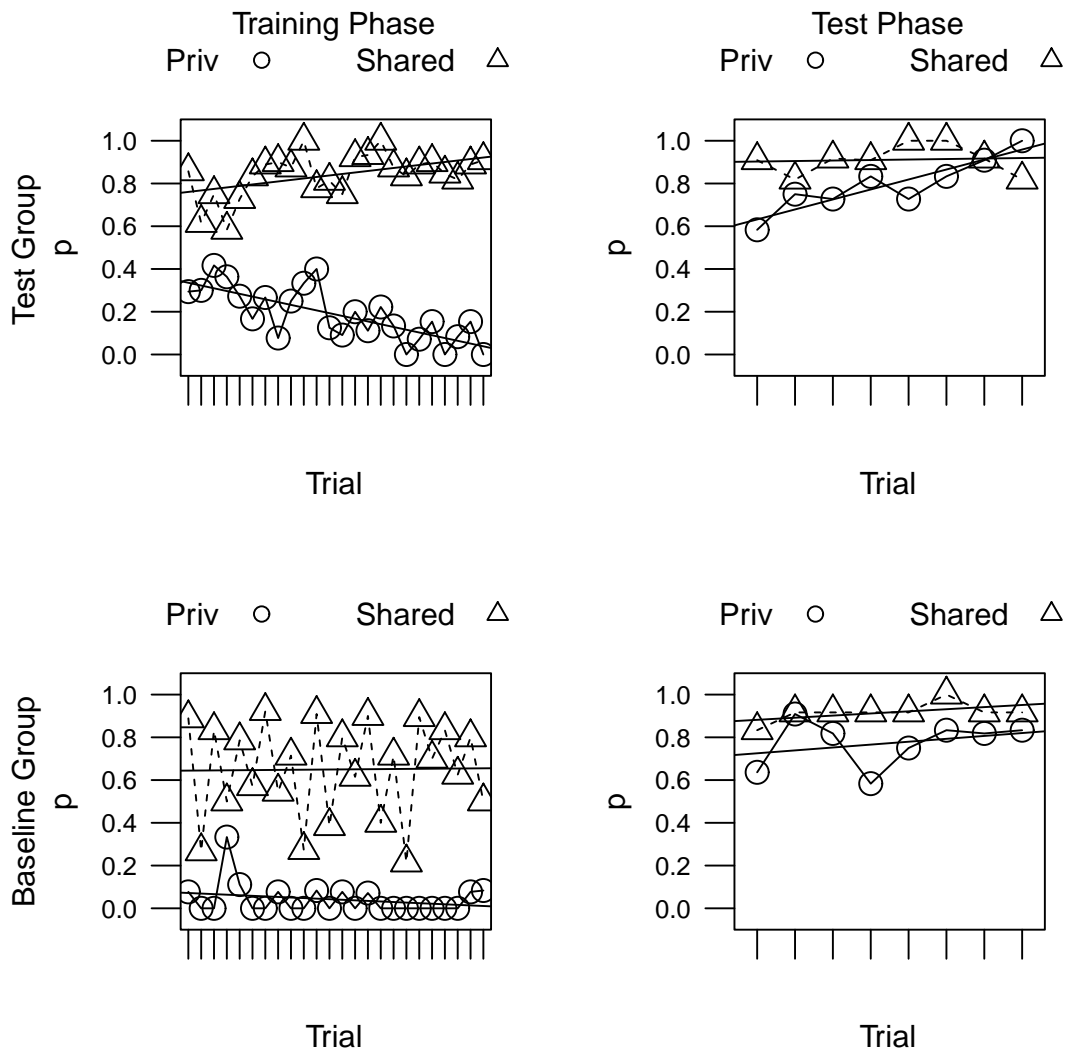Table 2.3: Inferential statistics for modifier use in the test phase, first for Order as a mean-centered variable (Ground refers to the effect at the midpoint of the experiment), and second for Order (first trial coded as 0 so that Ground is interpreted as the effect at the start of the phase).

Factor	Initial Effects			Centered Effects		
	Estimate	$\chi^2_{(1)}$	p	Estimate	$\chi^2_{(1)}$	p
Ground (G)	3.029	13.993	$< 0.001^{**}$	2.111	17.877	$< 0.001^{**}$
Baseline (B)	-0.021	0.004	0.985	0.066	0.004	0.949
Order (O)	0.284	6.626	0.010^*	”	”	”
G:B	1.317	0.587	0.444	-1.418	1.742	0.187
G:O	-0.262	1.214	0.271	”	”	”
B:O	0.025	0.013	0.911	”	”	”
G:B:O	-0.781	2.858	$0.091^\dagger$	”	”	”

$^\dagger p < 0.10$; $^* p < .05$; $^{**} p < .01$.

sponse (as an indicator of confusion) from the listener by editing in missing information after the fact. One way of capturing this is to consider whether speakers used post-nominal modifiers (e.g. "the bag that is small") more often in the privileged condition. In the training phase, post-nominal modifiers were relatively rare (5.2% of responses); in the test phase they remained at this low rate for the shared condition (5.9%) but are significantly higher in the privileged condition (27.2%; $\chi^2_{(1)} = 34.772, p < 0.001$, refer to Table 2.4 and Figure 2.5). As can be seen in Table 2.5 and Figure 2.6, when only pre-nominal modifiers are considered, the magnitude of the ground effect increases substantially. The significant ($\chi^2_{(1)} = 19.672, p < 0.001$) first-trial difference expands to 54% (90.9% in the shared condition, 37.5% in the privileged condition). These effects persist into the middle of the test phase ($\chi^2_{(1)} = 22.583, p < 0.001$) and are not moderated by a significant interaction with trial order ($\chi^2_{(1)} = 2.663, p = 0.103$). Had the interaction occurred, it would have suggested that the effect moderated over time, although this would still support an interpretation in line with monitoring and adjustment. If the test phase were longer, it is almost certain that this would be the case as speakers would be expected to adapt fully to their new circumstances. An alternate argument could

Table 2.4: Inferential statistics for post-modifier use.

Factor	Training Phase			Test Phase		
	Estimate	$\chi^2_{(1)}$	p	Estimate	$\chi^2_{(1)}$	p
Ground (G)	0.882	5.067	0.024*	-2.001	34.772	< 0.001**
Baseline (B)				-0.530	1.43	0.232
Order (O)	-0.016	0.286	0.593	-0.043	0.207	0.649
G:B				0.010	0.002	0.989
G:O	0.055	0.888	0.346	0.082	0.222	0.638
B:O				-0.037	0.048	0.826
G:B:O				-0.143	0.172	0.678

$^\dagger p < 0.10$; $^* p < .05$; $^{**} p < .01$.

Table 2.5: Inferential statistics for modifier use in the test phase with utterances that used post-modifiers removed, first for Order as a mean-centered variable so that the Ground refers to the effect at the midpoint of the experiment, and second for Order coded with the first trial as 0 so that Ground is interpreted as the effect at the start of the phase.

Factor	Initial Effects			Centered Effects		
	Estimate	$\chi^2_{(1)}$	p	Estimate	$\chi^2_{(1)}$	p
Ground (G)	3.839	19.672	< 0.001**	2.430	22.583	< 0.001**
Baseline (B)	0.198	0.028	0.866	0.426	0.181	0.671
Order (O)	0.269	5.292	0.021*	”	”	”
G:B	0.595	0.105	0.746	-1.574	2.105	0.147
G:O	-0.403	2.663	0.103	”	”	”
B:O	0.065	0.080	0.777	”	”	”
G:B:O	-0.620	1.667	0.197	”	”	”

$^\dagger p < 0.10$; $^* p < .05$; $^{**} p < .01$.

be made that participants were initially confused by the change in perspective, however in that case one would expect speakers to have more swiftly adjusted to the new perspective.

Speech onset time (truncated to the 97.5th percentile) in the test phase, as a measure of planning, was also considered (Figure 2.7 and Table 2.6). The baseline condition was initially faster ($\chi^2_{(1)} = 4.720, p = 0.03$) and was on average faster throughout the phase ($\chi^2_{(1)} = 8.062, p = 0.004$). It's probable that speakers planned their
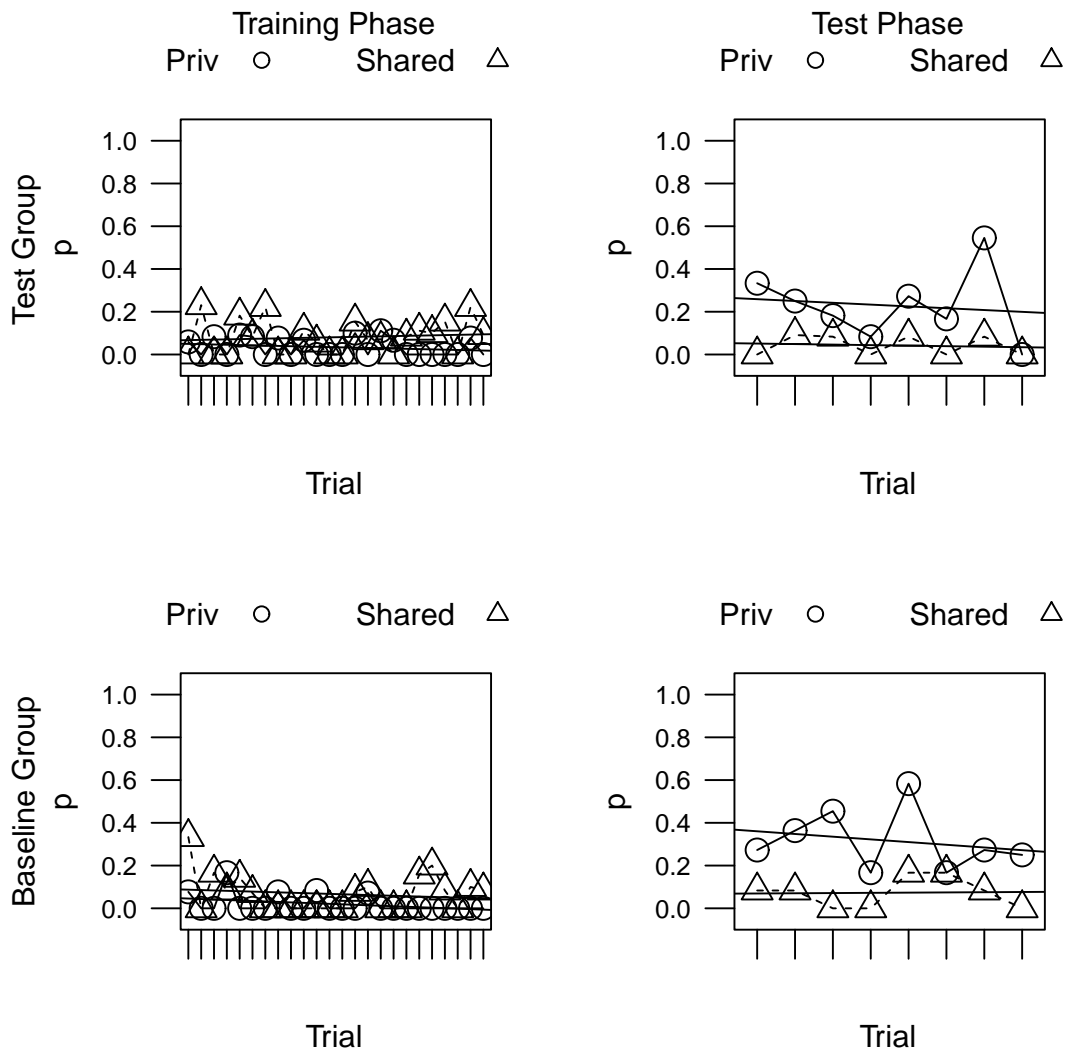
Figure 2.5: Proportion of trials in which the director used a size modifier in the post-nominal position.
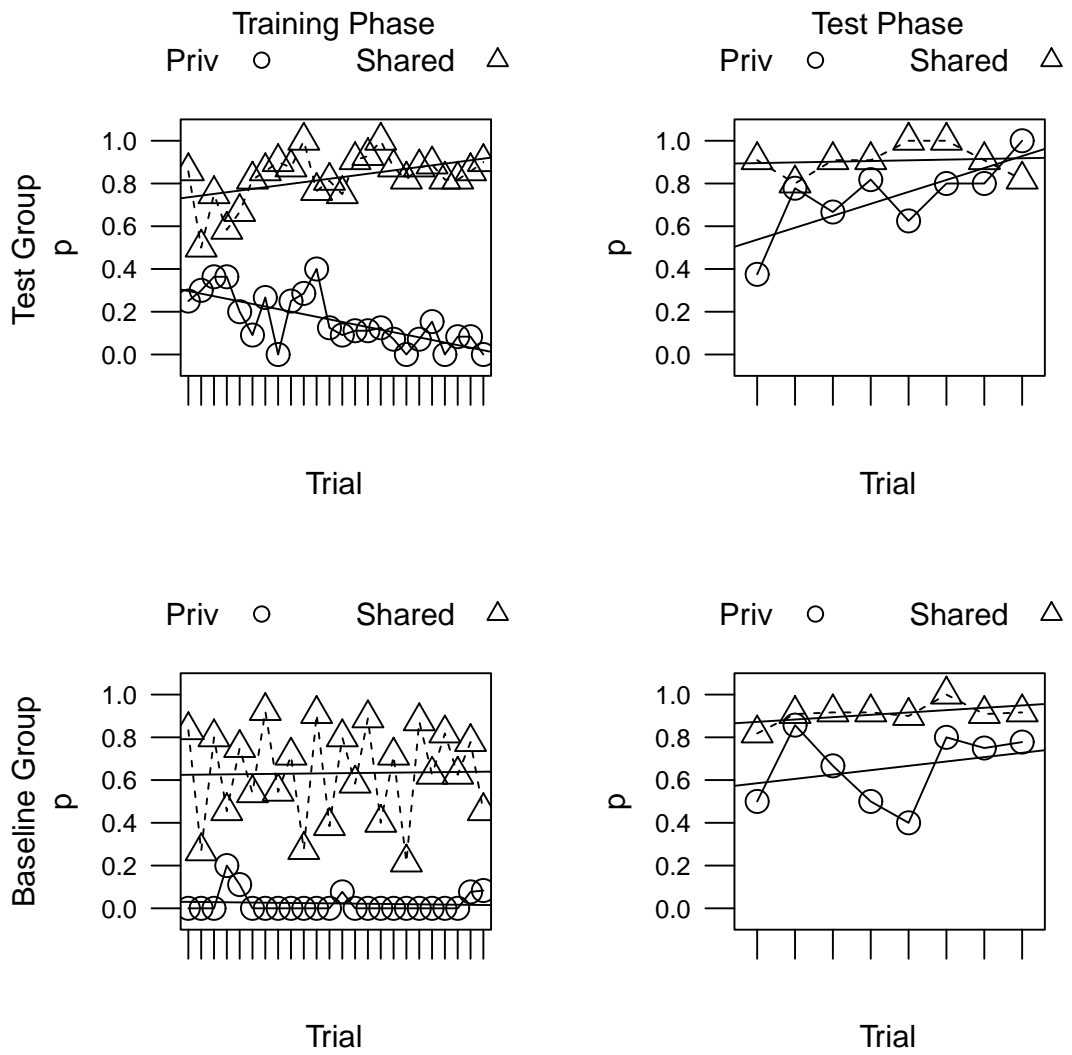
Figure 2.6: Proportion of trials in which the director used a size modifier in the pre-nominal position.

Table 2.6: Inferential statistics for speech onset time in the test phase, first for Order as a mean-centered variable so that the Ground refers to the effect at the midpoint of the experiment, and second for Order coded with the first trial as 0 so that Ground is interpreted as the effect at the start of the phase.

Factor	Initial Effects			Centered Effects		
	Estimate	$\chi^2_{(1)}$	p	Estimate	$\chi^2_{(1)}$	p
Ground (G)	-0.016	0.271	0.603	-0.057	13.248	$< 0.001^{**}$
Baseline (B)	0.091	4.720	0.030^*	0.100	8.062	0.004^{**}
Order (O)	0.003	0.107	0.743	"	"	"
G:B	-0.023	0.155	0.694	-0.035	1.236	0.266
G:O	-0.012	2.618	0.106	"	"	"
B:O	0.003	0.159	0.691	"	"	"
G:B:O	-0.003	0.052	0.819	"	"	"

$^\dagger p < 0.10$; $^* p < .05$; $^{**} p < .01$.

utterances more quickly because of the relative ease of the task in the baseline condition due to the lack of competitors in the privileged condition during training. Onset time was also slightly faster for the shared condition than in the privileged condition $(\chi^2_{(1)} = 13.248)$ over the course of the test phase. The marginal interaction with trial order $(\chi^2_{(1)} = 2.618, p = 0.106)$ suggests there may have been a speeding up over the test phase as there was no speed advantage at the start of the phase $(\chi^2_{(1)} = 0.107, p = 0.743)$. Had the test phase been longer, there may have been more conclusive evidence of privileged information affecting production time.

Finally, the eyetracking data were analyzed for the critical test phase, non-baseline condition. Refer to Table 2.7 and Figure 2.8, gaze was analyzed using an empirical logit transformation. Where necessary, references to empirical logits have been converted to proportions for ease of interpretation. In support of our hypothesis, there was a significant interaction between ground and trial order $(\chi^2_{(1)} = 28.803, p = 0.042)$. Very little overt attention was paid to the privileged corners at the start of the test phase prior to speech onset (less than 1% of the time). However, by the end of the test phase
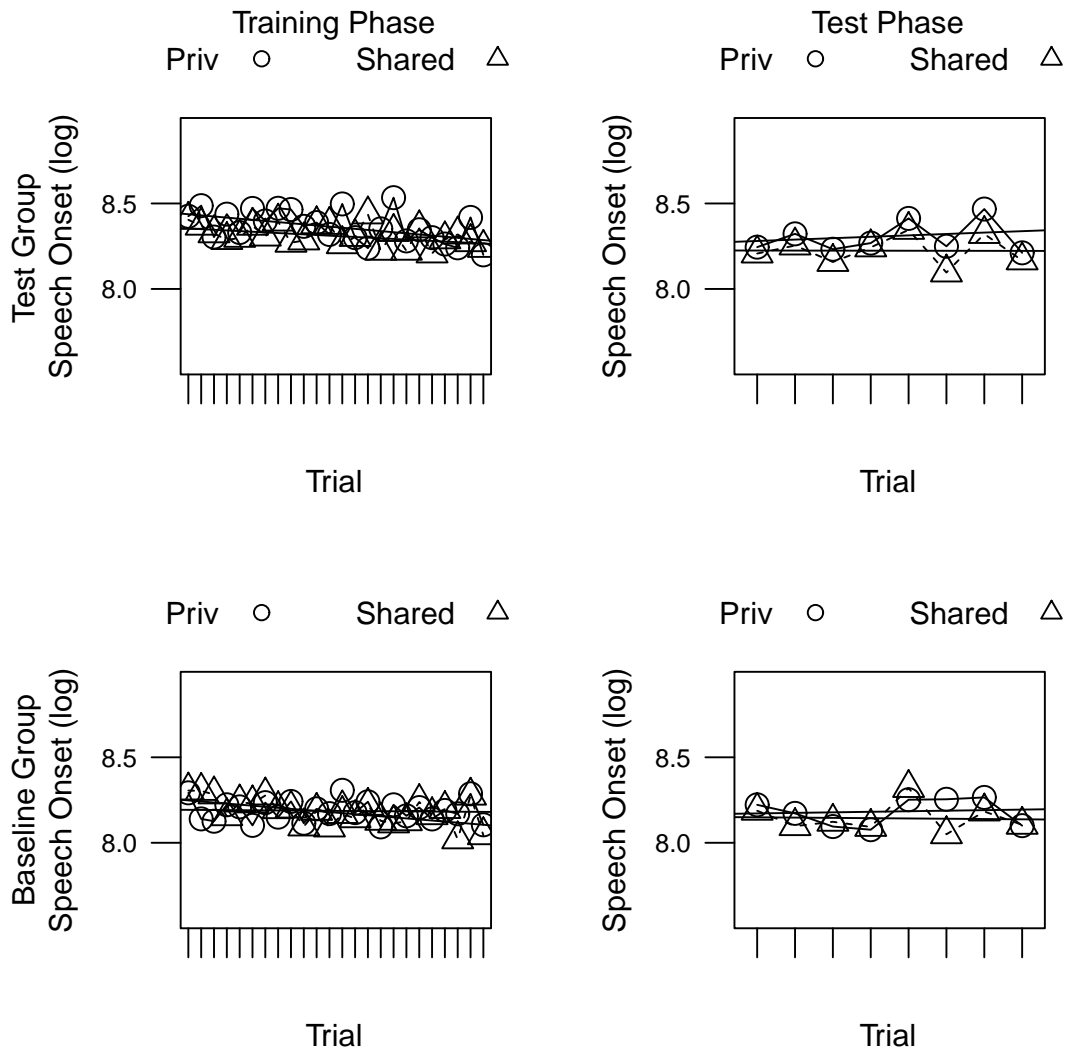
Figure 2.7: Speech onset time expressed on a logarithmic scale.

42
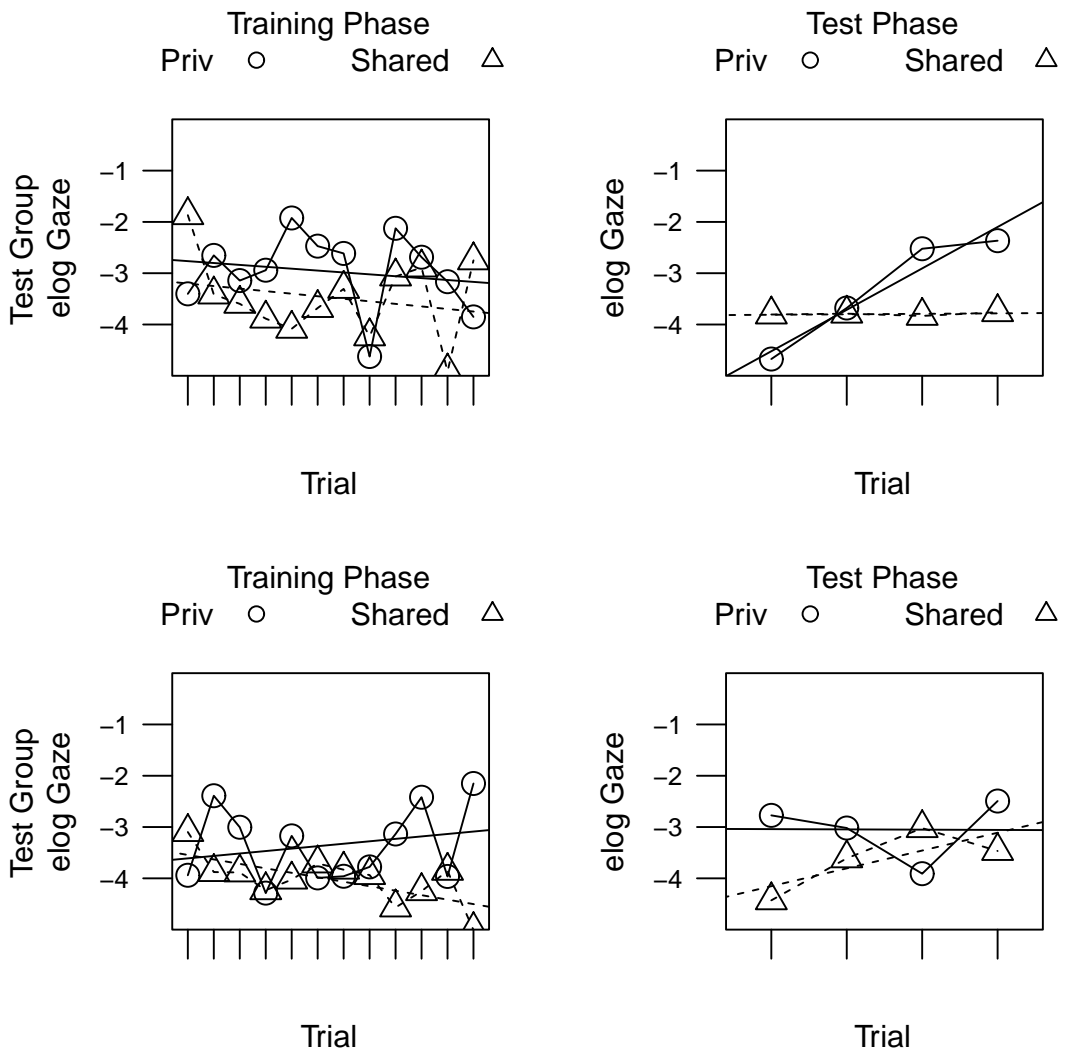
Figure 2.8: Descriptives for proportion of time spent fixating objects in the privileged corners of the display, in terms of empirical logits. For reference, $elog = -4$ is close to 1%, $elog = 0$ is 50%.

Table 2.7: Inferential statistics for proportion of time fixating the corners (privileged) in the display. Only includes test phase and test condition.

Factor	Initial Effects			Centered Effects		
	Estimate	$\chi^2_{(1)}$	p	Estimate	$\chi^2_{(1)}$	p
Ground (G)	0.684	3.346	0.067†	0.037	0.090	0.764
Order (O)	0.141	0.226	0.029*	"	"	"
G:O	-0.185	26.803	0.042*	"	"	"

$^\dagger p < 0.10$; $^* p < .05$; $^{**} p < .01$.

the corners were being fixated at a much higher rate (7.1%) in the privileged condition. Overt attention to the shared condition, by contrast, stayed flat throughout. This offers concrete evidence that speakers were not initially considering the corners of the display at the start of the test phase, and only began to allocate attention to those objects during the test phase, at the same time that their speech was also beginning to take those objects into account. An increasing amount of pre-nominal modification indicated that the initial utterance is accounting for the contrast rather than that information being encoded as a post-nominal adjustment.

2.4 Discussion

The major goals of this experiment were twofold. First, to determine whether participants facilitate audience design by managing interference from information that may be salient to them, but not available to their partners, by suppressing their attention to that information. This was achieved in the first phase of the experiment by having speakers engage in a series of training trials that acclimated them to the task, during which they were made aware that privileged competitors would always be consistently placed in the corners of the display. The results were mixed. Although the speech data indicated that speakers were showing a near-ceiling effect of common ground status by the end of the training, there was only a marginal effect of a decline in overt attention

devoted to the privileged objects (as measured by gaze duration). On the other hand, overt attention to the privileged regions of the display was extremely low throughout the experiment (until a notable bump in the test phase mentioned below), and time spent viewing those regions was comparable between the test and baseline group, despite the absence of competitors in the privileged region in the baseline group. It's possible that viewing of the corners was demonstrating a floor effect.

The second goal was to determine whether this suppression of the use of privileged information persisted despite a sudden shift in perspective that brought the previously suppressed information back into prominence. That is, do speakers quickly and easily adjust to changes, and what constitutes common ground in a way that implies an early impact of ground on utterance planning in these circumstances? As predicted, the data showed that speakers were significantly less likely to incorporate formerly privileged information into their utterances, especially when post-nominal modification/repairs are removed from the analysis. The greatly increased likelihood of post-nominal repairs in the test phase could have reflected a response by the speaker to confusion, but could also have been a result of late-fixation to the competitor after speech onset (Brown-Schmidt & Tanenhaus, 2006; Brown-Schmidt & Konopka, 2008). It is difficult to determine the difference given these data. In the test phase, gaze to the privileged regions during utterance planning was initially insensitive to the presence of a formerly privileged competitor, with a sharp rise in attention to privileged objects during planning as the test phase progresses. This, however, was not accompanied by a drop in the use of post-nominal correction, which might make it more likely that speakers are responding to listener cues, given that the listeners would hesitate when given a non-definitive description.

It is important to consider the external validity of these results given the un-

usual nature of the experiment. This setup is less unreasonable than it might seem at first glance. Traditional studies of common ground have the location and nature of the privileged information randomized between trials. In effect, speakers have to prepare their utterances from scratch in regards to common ground from trial to trial. That type of design thwarts the possibility of allowing a speaker to attenuate their attention to information that is deleterious to audience design because the context must be considered in its entirety in each new trial. A strategy of adjusting and re-allocating loci of attention is not possible in such a situation, whereas information can be fore-grounded and backgrounded in working memory based on the accumulation of evidence and the changing focus of attention in dialogue. In live conversation, the body of what constitutes common ground is grown more organically and builds on itself, opening the possibility of a re-weighting of the salience of different channels of information relevant to the audience design process. Previous studies (e.g. Roßnagel, 2000) that involved an accumulation of common ground over the course of an experiment, focused on different issues than the allocation of attention. In a complimentary fashion, results such as the ironic effect seen in Wardlow Lane et al. (2006) show how circumstances that increase the salience of privileged information lead to its increased use despite the demands of audience design.

Another important factor to note is that the size of the effect of ground in the training phase was substantially influenced by how many trials had elapsed. Both the use of the appropriate modifier in the shared condition and its non-use in the privileged condition was dependent on practice. In this case, speakers had to learn to associate the location of the object with its status with respect to common ground. These results could be used as an argument for either substantial practice prior to the commencement of an experiment, or more experimental trials in general, in order to maximize effects of

ground. Even in experiments in which common ground is signaled in a more naturalistic fashion, speakers still need an opportunity to adjust. For example, in the current experiment addressees moved to looking at the same display as speakers. Yet, the speaker still required time to adjust to the new common ground.

These results fit in with previous results, such as those of Brennan and Clark (1996) with partner swapping, that demonstrate speakers need time to adapt to new circumstances and that a speaker's deployment of new information about what is included in common ground is not immediate. While the current experiment focused on binding perspective information to physical location and entrainment in terms of how attention is allocated, the next experiment focuses on how entrainment in memory can lead to similar effects.

Chapter 3

Experiment II

3.1 Introduction

Audience design can be seen as a form of expertise focused on solving the problem of how to assign a name to a referent. Logan's Instance Theory of Automaticity (ITA; Logan, 1988) (which is an ancestor model of ITAM) offers one account of how this expertise is developed through the accumulation of instances in memory which store previously used referential descriptions (for a citation of ITA in language comprehension, see Rawson, 2010). The first time a speaker has to refer to an object, they are forced to engage in a relatively slow deliberative process that may include an explicit consideration of what is included in common ground. Subsequent encounters with the referent have the advantage of being able to rely on the memory traces formed by previous experience. This reliance on pre-established memory should increase with familiarity. This is because if decision making regarding what form of utterance to use is based on a race model, automatic retrieval should be increasingly able to supply a sufficient solution (and thus always "win" the race) as experience is accumulated. The use of memory routines also offers an opportunity to differentiate which aspects of audience design are more subject

to automatic processing, such as the retrieval of names, as opposed to other components which may be more controlled like the incorporation of common ground. One way to determine the degree to which this is the case would be to give speakers an opportunity to develop routines and to see how the use of those routines responds to changes in the context within which they are deployed.

Speakers tend to calibrate their speech to meet the needs of their listener, incorporating neither too much nor too little information into an utterance than is needed by their addressee to identify it (Grice, 1975). However, speakers do depart from this norm under certain conditions. For example, speakers overspecify their descriptions, providing unneeded privileged information, when they are under cognitive load (Horton & Keysar, 1996) and when that privileged information's salience is heightened (Wardlow Lane et al., 2006). Overspecification is far more common than underspecifying a referent (Ferreira, Slevc, & Rogers, 2005). One goal of the present study is to develop a situation in which underspecified referents are more likely to be observed. One way to achieve this would be to first have speakers routinize a description in a particular context, and then alter the context so that the established routine would offer insufficient information to allow the addressee to reliably identify the referent.

The current experiment seeks to achieve a greater understanding of the different contexts that lead speakers to make these kinds of errors. The first factor is to setup a situation that exploits speakers' use of routines in order to lead them into either overspecifying or underspecifying their utterances. To elicit overspecification, we allow speakers to repeatedly refer to a concrete, easily named, object in the presence of a competitor, and then see whether they preferentially continue to use that routinized description when the competitor is removed. Previous work by Wilkes-Gibbs and Clark (1992) and Krauss and Weinheimer (1966) has demonstrated that when speakers repeat-

edly refer to abstract, initially difficult to describe objects, their descriptions become shorter and more concise. If speakers then have to refer to the object with a new addressee who is unfamiliar with the object, they may preferentially rely on their routinized description even though it may not be adequate for a naïve addressee. Regardless of whether they are referring to concrete or more abstract objects, speakers should be seen to be preferentially relying on the use of the descriptions they had utilized previously as opposed to treating old referents as if they were novel.

Wilkes-Gibbs and Clark (1992) had participants perform a similar task. However, in their experiment the allowance of unlimited feedback between interlocutors may have allowed speakers to revise their utterances and mask the appearance of underspecified utterances. It is necessary then to manipulate the availability of verbal feedback. Not only because direct feedback may lead speakers to revise their utterances, but because even the *potential* for feedback may have consequences for how a speaker plans their utterances. The potential for feedback may allow speakers to more freely rely on routines because they know their addressees can request clarification if needed. Thus when speakers are unable to rely on the availability of feedback they may take more time to plan their utterances to make sure they are adequate, and delay the onset of speech. There is evidence that speakers will respond to the demands of particular situations by adjusting their decision criterion for settling on a particular utterance (Roßnagel, 2000). Thus, lacking feedback may make speakers feel more accountable for the quality of their utterances and they may attempt to rely more on a slower deliberative process to produce their utterance. The availability of feedback makes misspecification potentially less costly, and speakers may then adopt a more liberal criterion.

If speakers are relying on routines, speakers should adapt less to new addressees. This is important for eliciting underspecified utterances with abstract ref-

erents, as speakers should maintain the use of more concise descriptions developed with their original partners that may be informationally impoverished for a novel partner. Likewise, speakers should retain the richer utterances developed for concrete referents in the presence of a competitor during training, despite the absence of that context in the test phase. At first glance, these hypotheses appear to conflict with the conclusions of Brennan and Clark (1996), however the assertion of insensitivity to a naïve partner is actually consistent with their data because speakers did not adapt immediately after the partner switch. They only show an effect of adaptation to a new partner when they average over all trials in their test block; their data show that this adaption happened over time and was not evident at the beginning of their test block. There may also be a relationship between feedback and speakers' adaptation, as seen in the first experiment where speakers had a tendency to add content to initially inadequate descriptions.

3.2 Method

3.2.1 Participants

The experiment involved 80 participants (47 females and 33 males), who formed 16 triads and 16 dyads, drawn from the undergraduate population of the University of California, Riverside. The experimenter randomly assigned members of each dyad or triad to their respective roles: Director (speaker), Matcher A (old addressee), and Matcher B (new addressee).

3.2.2 Apparatus

The apparatus was the same as it was in the previous experiment.

3.2.3 Materials

The experimental stimuli consisted of 42 sets of pictures. Each set consisted of a target object, a competitor object, and four unrelated filler objects (one of which replaced the competitor during the test phase of the experiment). Two of the sets were used for practice trials that were not included in the analyses. Of the 40 experimental sets, 16 included a concrete target, 16 had an abstract target (these objects have no conventional names), and 8 sets had concrete targets that did not have a competitor. Each of the stimulus images were 200x200 pixel bitmaps with the object pictures set on a black background. Descriptions of the pictures used as concrete targets are given in Appendix B.

3.2.4 Procedure

The experiment began with a pre-recorded audiovisual presentation that introduced participants to the task. An example display containing five objects was shown to the participants, and the director was informed that her task was to describe one of these, the target, such that the matcher can identify it. The director was informed that she was not allowed to identify the target by mentioning the number of the space it appeared in or by otherwise indicating its location. The matchers' task was described as simply to select the target with the mouse. They were told that in the "Feedback" trials, the matcher would be able to start moving the cursor immediately, and that participants would be able to freely converse. In the "No Feedback" trials, matchers were told that they were not allowed to talk to the director, nor to move the mouse until the director had finished her description. The director was told to indicate that she had finished the description by pressing a button on a response pad. Triad participants were

informed that there would only be one matcher present in the room for a given block, and all participants were also reminded that the matcher who is outside of the room is not able to hear or see what is going on inside the room. They were also told to take care to remember who is matcher A and who is matcher B. In the sessions involving triads, the matcher not participating in the current block waited out of earshot in an adjacent room until it was their turn to participate.

At the beginning of a trial, a display appeared on the screen showing the numbers 1 to 5 arranged in the locations where the pictures would eventually appear. After viewing this screen for 1000 ms, speakers heard a pre-recorded voice through headphones that announced the location of the target. 2000 ms after this, the numbers were replaced with five objects. In this way, speakers gaze would tend to be on the target item when the objects appeared, and the number of shifts away from the target could be taken as a measure of the use of perceptually co-present information.

On "No Feedback" trials, speakers gave their description and pressed a button on a gamepad controller to indicate they had completed the description to their satisfaction. Prior to the button event, the addressee's mouse cursor was frozen in the middle of the screen. The addressee was only able to move the mouse after the button had been pressed. Addressees were not allowed to talk to the speaker on these trials. On "Feedback" trials, addressees could freely interact with the speaker. Also, the speaker did not need to press the button to indicate completion, and the listener's mouse cursor could be moved from the onset of the main display.

The experiment was divided into four blocks of trials, each of which in turn was divided into a training and a test phase. An example block of trials is given schematically in Figure 3.1. In the training phase of each block, speakers referred to two concrete and two abstract targets presented five times each in a random order (for a total of 25 trials

per training block), thus establishing descriptions for these referents. In addition, they referred to a filler target during five additional filler trials during the training. Just prior to the test phase, there was a transition phase during which the director was presented with a 30 second countdown until the start of the test phase, which gave the experimenter the opportunity to swap Matcher A for Matcher B and vice versa as necessary when there were multiple matchers. The countdown also served to give the directors in the non-partner switching condition an equivalent amount of downtime between the training and test phases. At the end of the countdown, the participants were presented with a screen reminding them as to whether or not feedback was allowed in the subsequent phase.

After the transition events, the test phase began. In this phase of each block, speakers once again referred to the two old concrete and two old abstract referents, in addition to two new concrete and two new abstract referents. There were also two filler displays. The abstract, concrete and filler displays appeared in a random order within their block. Unlike the training phase, when concrete referents appeared as targets, in the test phase they appeared without the same-category competitor.

3.2.5 Design

The experiment included factors of phase (training versus test), item conventionality (concrete versus abstract), item novelty (was it new to the test phase), partner (switched between phases or not), and feedback (allowed during test or not). The test phase was the critical phase and the training is ignored in the analyses unless specifically noted. Feedback was manipulated within-subjects, however partner switching was between subjects. Analyses for the conventionality conditions were conducted separately
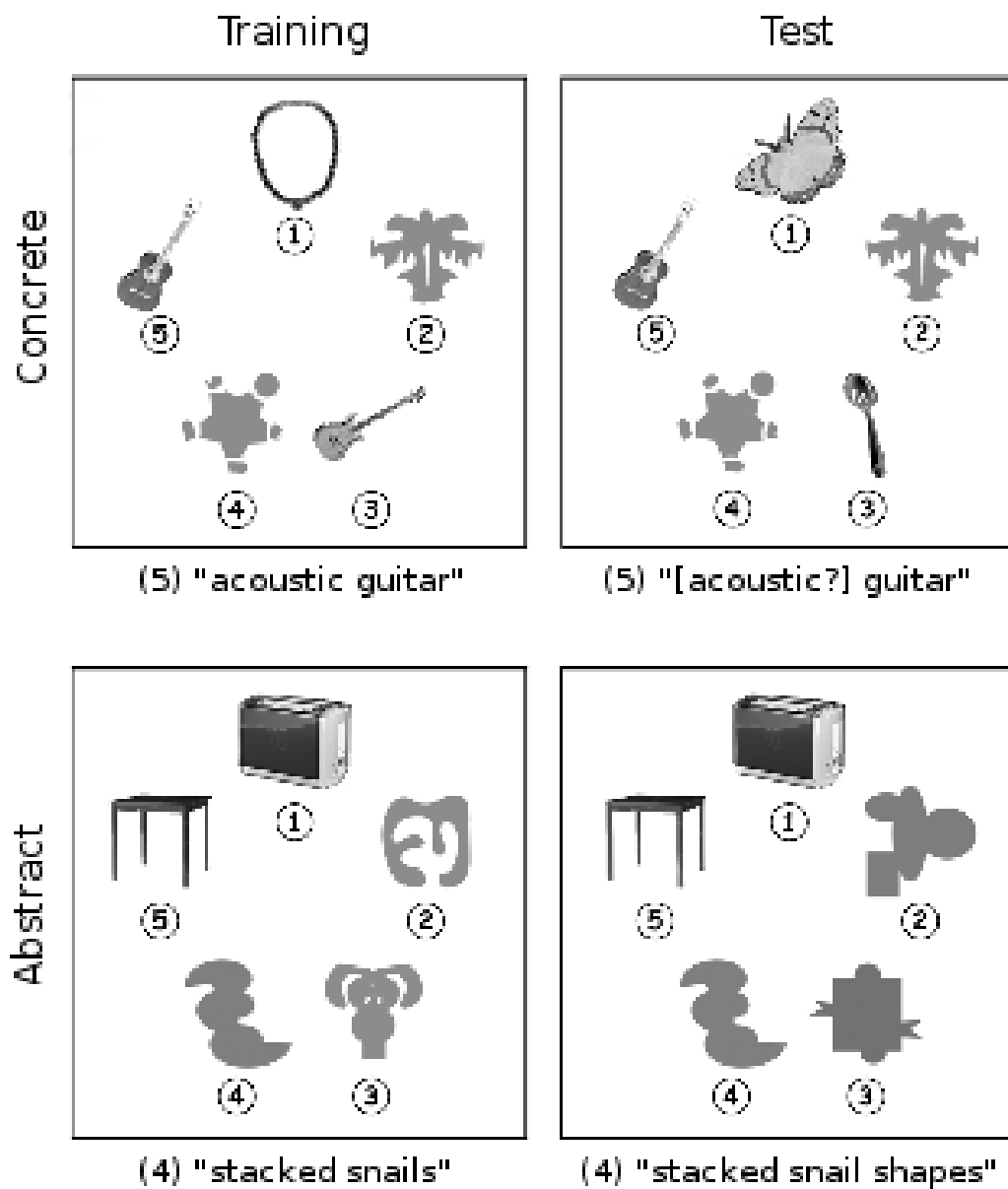
Figure 3.1: Example displays for the two experimental phases for concrete and abstract referents. Below each display is a number identifying the target (which the director hears at the start of the trial) and a sample utterance of what might be said in each condition.

as they were meant to explore fundamentally different hypotheses. The dependent variables were accuracy (whether the addressee chose the correct target), speech onset time, utterance length (number of words in the utterance), and gaze duration (in number of frames on the area of interest).

3.2.6 Analysis

The director's speech was recorded, transcribed, and coded for accuracy, speech onset, and utterance length by the experimenters, who were blind to the condition. The eye-tracking data were coded as number of frames (sampled at 60 hz) that the director spent look at each object prior to speech onset. Analyses were conducted separately for the abstract and concrete items, since as noted above, those items sets were designed to test different questions regarding underspecifiation and overspecification, respectively. In addition the speech in the abstract condition was further coded by the content of the speech in the test phase as compared to the corresponding last trial of the training phase (only for items repeated between phases). Utterances that were the same between phases were coded as "pure repetitions," utterances that were initially the same but added words were coded as "elaborated repetitions," utterances that were shorter in the test phases were coded as "reductions," and finally utterances that had different wording between phases were classified as "reconceptualizations."

The analyses were conducted using the same modeling techniques described in the analysis of the first experiment.

3.3 Results

3.3.1 Concrete Referents

The analysis of the concrete items focuses heavily on the question of whether speakers tended to overspecify referents in a novel situation if they had prior experience referring to the same referent in a different context. Before considering the results, it is important to note that speakers would only have had the opportunity to overspecify during the test phase if they had used specific terms during the training phase. Even though all training trials for concrete referents included a pair of objects from the same category, a target typical of the lexical category (e.g., an average-looking candle) as well as a competitor that was less typical (e.g., a melted candle)–on those trials speakers sometimes used only a bare noun to identify the target, e.g., calling the candle that was not melted "the candle". Because such cases created no opportunity for overspecification at test, we removed any test trial for which the speaker had used the bare noun for the target on the training trial just prior to the test trial. This led to the elimination of 43 out of 480 trials (9%). We also eliminated test trials where speakers miscategorized the target (e.g, calling a red candle the "gas can" or the "red cylinder") or otherwise described the target in a way that would not apply to the competitor (e.g., calling the adult gorilla "King Kong"). This resulted in the removal of an additional 11 trials (2.3%). The final data set therefore contained a total of 427 out of 480 possible observations (89%). Summary statistics are listed in Tables 3.1 and 3.2. Visualizations of these data are shown in Figures 3.2 and 3.3.

Overall, on the test trials, where speakers described old targets in the absence of any competitor, they overspecified the referent about 68.9% of the time; e.g., they called the typical candle "the unmelted candle". This was much higher than the 8.2% baseline

level of overspecification for new targets (Novelty: $\chi^2_{(1)} = 30.54, p < .001$). If speakers took the addressees' knowledge into account, then they should have been less likely to overspecify the referent when speaking to a new partner. However, overspecification rate seemed to depend only on whether the speaker had been trained on a referent, not the identity of the addressee. With the old partner, speakers overspecified old referents at a rate (76%) that was about 9.5 times higher than baseline (8%), whereas with the new partner, they did so at a rate (62%) that was about seven times higher than baseline (9%); however, there was little evidence that the increase relative to baseline differed across addressee (Addressee-by-Novelty interaction: $\chi^2_{(1)} = .84, p = .360$). Whether or not speakers overspecified did not appear sensitive to the availability of feedback ($ps \leq .438$ for all main effects and interactions associated with this factor).

Table 3.1: Inferential statistics for concrete referents based on over-specification rate (use of an adjective in the test condition) and speech onset time.

Factor	Over-specification Rate		Speech Onset	
	$\chi^2_{(1)}$	p	$\chi^2_{(1)}$	p
Addressee (A)	0.27	0.604	1.77	0.183
Novelty (N)	29.88	< 0.001**	6.73	0.009**
Feedback (F)	0.25	0.618	0.15	0.694
A:N	0.84	0.360	1.69	0.194
A:F	0.01	0.937	0.27	0.603
N:F	0.34	0.558	0.21	0.648
A:N:F	0.16	0.693	1.90	0.168

$^{\dagger}p < 0.10$; *$p < .05$; **$p < .01$.

Consistent with the absence of robust audience design effects in the production measure, none of the processing measures showed any evidence for audience design. The time that speakers spent planning (speech onset was trimmed at the 97.5th percentile, 2645 ms) was influenced only by the novelty of the referent, with speech starting 172 ms faster on average before an old than a new referent (Novelty: $\chi^2_{(1)} = 6.73, p = .009$).
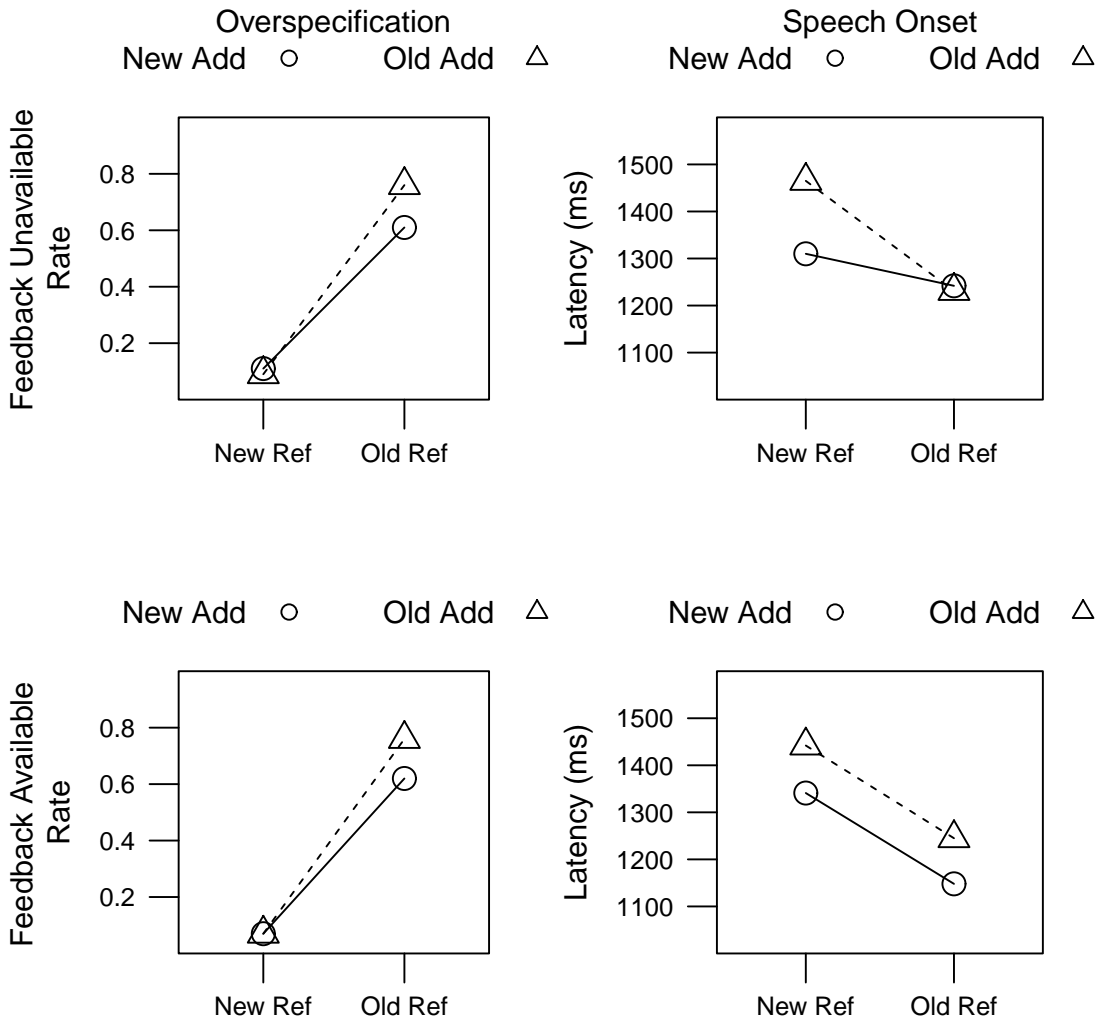
Figure 3.2: The columns correspond to the two production measures, over-specification rate and speech onset, for the concrete items.
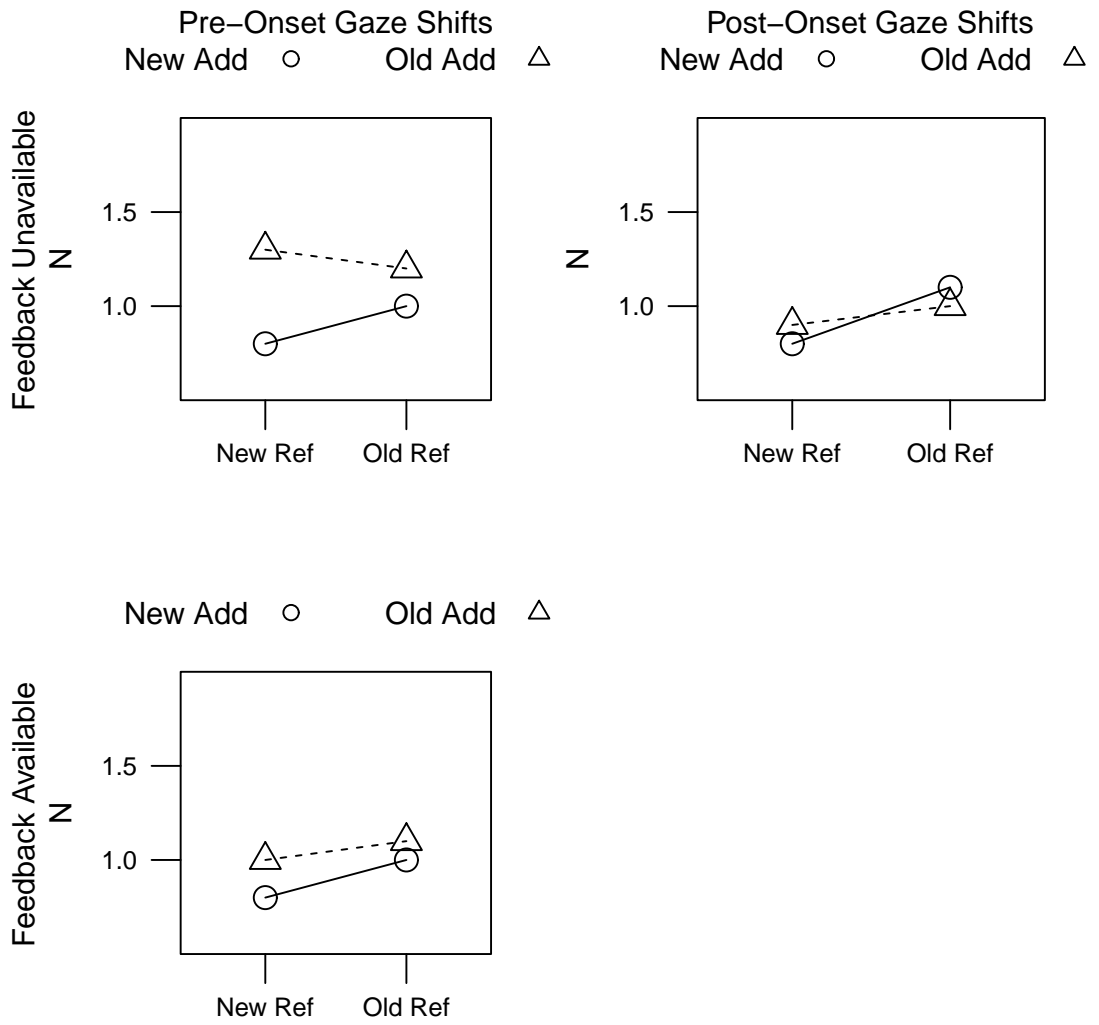
Figure 3.3: Number of pre- and post-onset gaze shifts for concrete trials.

Table 3.2: Inferential statistics for concrete referents using gaze measures.

	Pre-Onset		Post-Onset	
Factor	$\chi^2_{(1)}$	p	$\chi^2_{(1)}$	p
Addressee (A)	2.49	0.115	0.13	0.721
Novelty (N)	1.21	0.271	1.29	0.257
Feedback (F)	0.02	0.891		
A:N	0.33	0.567	1.13	0.287
A:F	0.26	0.613		
N:F	1.21	0.271		
A:N:F	0.01	0.951		

$^\dagger p < 0.10$; $^* p < .05$; $^{**} p < .01$.

There were no significant effects related to pre-onset or post-onset gaze shifts (which were truncated at the 97.5th percentile, which was 4 for both). However, it is possible that the gaze shift measures were near floor, since there does not seem to be even a trend toward fewer gaze shifts in the old referent condition, as would be expected.

There was little evidence that speakers tailored their descriptions of concrete referents to the addressee's informational needs. This fits with our previous assessment of Brennan and Clark (1996), in which adaption was not seen to be immediate, but rather happened over the course of several trials. In our experiment, there were only references to two old concrete referents in each test block, and thus little opportunity for speakers to get the feedback from the listener that might have caused them to adjust their descriptions. Given these considerations, our current results are fully consistent with those of Brennan and Clark. Moreover, our results show that speakers are prone to overspecify referents even when the overly specific terms are highly anomalous in a neutral context, such as calling a prototypical candle "the unmelted candle."

One question is whether speakers' overspecifications actually impaired comprehension, given previous suggestions that they can sometimes facilitate visual search (Arts, Maes, Noordman, & Jansen, 2011; Mangold & Pobel, 1988). To test this, we

examined listeners response latency (truncated to 6473 ms, the 97.5th percentile) as a function of whether or not speakers overspecified the referent. It only made sense to examine response times for descriptions of old referents in the feedback-available condition, since in the feedback-unavailable condition listeners were only able to move the mouse cursor after speakers pressed a button; such responses were therefore done "offline." We measured the latency to click the target from the onset of the noun (e.g., "candle" in "unmelted candle"). Listeners were indeed about 202 ms slower overall to click the target after hearing an overspecified description relative to a bare noun (4423 vs. 4221 ms, respectively; $\chi^2_{(1)} = 5.79, p = .016$). Surprisingly, there was no evidence that overspecified descriptions impaired comprehension more when addressees were unaware of the established description (161 ms vs. 243 ms penalty for new vs. old addressees; $\chi^2_{(1)} = .09, p = .759$). It is not clear why this is the case; one possibility is that even though informed addressees had heard the description "unmelted candle" multiple times, they had always heard it with contextual support, and perhaps found the absence of such support anomalous.

3.3.2   Abstract Referents

For abstract referents, if speakers are egocentric and rely wholly on the abbreviated descriptions they have established for old referents, they risk producing underspecified descriptions that could create a misunderstanding or make interpretation difficult for the addressee. Our primary measure of underspecification, then, was the effect of Novelty: that is, the difference in description length for old versus new referents. A completely egocentric speaker would use as few words to describe an old referent for a new addressee as they would for an old one; conversely, "ideal speakers" who are perfectly sensitive to their addressees informational needs should describe an old referent

to a new addressee using just as many words as they would have used had they been describing a new referent to that same addressee.

The length of descriptions (truncated to the 97.5th percentile which was 26 words) that speakers gave varied with their partner's knowledge (Addressee-by-Novelty interaction, $\chi^2_{(1)} = 36.16, p < .001$). There was little evidence that speakers underspecified old referents: they used 55% longer descriptions when talking about these referents to new addressees compared to old addressees (8.3 versus 5.4 words; $\chi^2_{(1)} = 9.21, p = .002$), replicating Wilkes-Gibbs and Clark (1992). There was still some evidence that speakers were influenced by their own knowledge, as the mean description lengths for new addressees and old referents (8.3 words) was, on average, about 11% shorter than that for new addressees and new referents (9.3 words), $\chi^2_{(1)} = 5.31, p = .021$. One surprising result was that the Addressee-by-Novelty interaction was driven not only by a partner effect for old referents, but also by a partner effect for new referents: speakers used about 20% more words when speaking about new referents to old versus new addressees (11.5 vs. 9.3 words; simple effect of Addressee, $\chi^2_{(1)} = 5.31, p = .021$).

The next question to consider is whether speakers managed to avoid underspecification through their own initiative, or because they were prompted to give longer descriptions by feedback from the new addressee. To address this question, we examine the effect of feedback availability. Speaker adaptation can perform self-prompted adaptations regardless of the availability of feedback, but other-prompted adaptation can only happen when addressees are free to offer feedback to act upon. Thus, the degree to which speakers are lengthening their description of old referents for new addresses when feedback is available, as compared to when it is not, should reflect whether these adaptations are other-prompted. Supporting the idea that this adjustment was self-

prompted, the three-way interaction was not reliable, $\chi^2_{(1)} = .01, p = .949$, with speakers lengthening their descriptions by about 61% (from 4.3 to 7.0 words) when feedback was available, and 51% when it was not (from 6.4 to 9.7 words). Although feedback availability did not seem to modulate speakers adjustments to addressees, it is worth noting that speakers were sensitive overall to feedback availability, giving descriptions that were about 35% longer when feedback was unavailable (9.9 vs. 7.3 words; main effect of Feedback Availability: $\chi^2_{(1)} = 6.51, p = .011$). The fact that speakers gave longer descriptions of old referents to new addressees when feedback was unavailable could only be explained as a self-prompted adaptation.

Speakers can and do spontaneously avoid underspecification by lengthening their descriptions of old referents for new addressees. A further critical question concerns whether speakers achieve this through allocating extra effort to planning, or through monitoring and adjustment. To answer this question, we turn to the on-line processing measures of speech onset latency and visual scanning. To the extent that speakers chose to include more information during the initial design of their descriptions, this extra planning should have delayed the onset of speaking relative to speaking to the old partner (truncated at 4216 ms, the 97.5th percentile). There was no evidence for this. Although the data showed the predicted interaction (Addressee-by-Novelty: $\chi^2_{(1)} = 10.40, p = .001$), unexpectedly, the partner effect was reliable only when speakers described new referents (2465 vs. 1954 ms for old and new addressees, respectively; $\chi^2_{(1)} = 4.08, p = .043$) but not when speakers described old ones (means = 1468 vs. 1346 ms for new vs. old addressees, respectively; $\chi^2_{(1)} = 2.44, p = .119$; see Figure 3.4). This finding, though unexpected, is consistent with the earlier finding that speakers gave longer descriptions of new referents to old (vs. new) addressees. It is intriguing that the longer descriptions that speakers gave for new referents and old addresses seem to have been at least partly

planned, whereas those given for old referents and new addressees seem to be wholly the result of monitoring and adjustment.

Further insight into the nature of these partner adaptations can potentially be found in the eye gaze data (number of gaze shifts truncated to the 97.5th percentile, 5 for pre-onset and 9 for post-onset). Speakers should rely more on physically co-present information (and thus, should show a greater rate of gaze shifts) when planning and monitoring descriptions of old referents for new addressees. This increase in the use of visual context could be reflected either in planning (pre-onset gaze shifts) or in monitoring (post-onset gaze shifts). Overall, speakers did scan less when they planned descriptions of old vs. new referents (.9 versus 1.6 pre-onset gaze shifts respectively, main effect of Novelty: $\chi^2_{(1)} = 7.56, p = .006$); however, there was no evidence that the scanning was partner-specific: although the Addressee-by-Novelty interaction was marginally reliable, $\chi^2_{(1)} = 2.97, p = .085$, there was no evidence that speakers relied more on physically co-present information when planning descriptions of old referents for new addressees (means of .95 and .95 for old and new partners respectively, $\chi^2_{(1)} = .06, p = .810$).

Table 3.3: Inferential statistics for abstract referents using production measures.

Factor	Word Count		Speech Onset	
	$\chi^2_{(1)}$	p	$\chi^2_{(1)}$	p
Addressee (A)	0.91	0.340	1.04	0.307
Novelty (N)	34.30	$< 0.001^{**}$	34.60	$< 0.001^{**}$
Feedback (F)	6.51	0.011^*	4.12	0.042^*
A:N	36.16	$< 0.001^{**}$	10.40	0.001^{**}
A:F	0.21	0.649	0.87	0.351
N:F	0.29	0.589	0.12	0.726
A:N:F	0.01	0.949	1.28	0.257

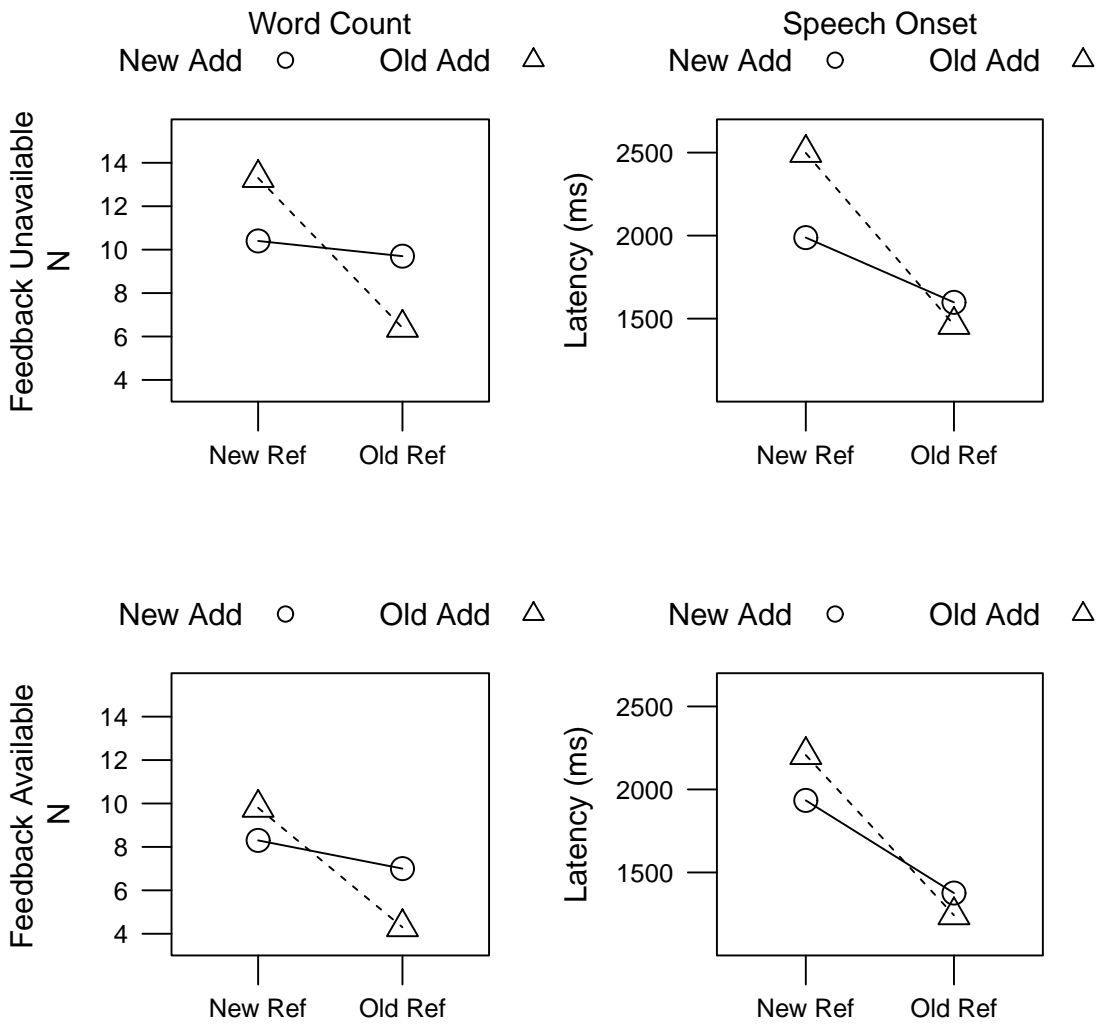$^\dagger p < 0.10$; $^* p < .05$; $^{**} p < .01$.

Figure 3.4: The columns correspond to the two production measures for the abstract items, word count (length of the description) and speech onset time.

Table 3.4: Abstract referents description analysis

Addressee	Feedback	Repetition Reduction	Elaborated Repetition	Reconceptualization
Old	Yes	73%	0%	27%
Old	No	55%	11%	34%
New	Yes	11%	33%	56%
New	No	13%	15%	72%

Table 3.5: Inferential statistics for abstract referents using gaze measures.

Factor	Pre-Onset $\chi^2_{(1)}$	p	Post-Onset $\chi^2_{(1)}$	p
Addressee (A)	0.86	0.354	0.45	0.501
Novelty (N)	7.56	0.006**	12.50	< 0.001**
Feedback (F)	0.76	0.383		
A:N	2.97	0.085†	7.59	0.006**
A:F	3.79	0.052†		
N:F	0.92	0.336		
A:N:F	2.93	0.087†		

$^\dagger p < 0.10$; $^*p < .05$; $^{**}p < .01$.

These considerations are partially qualified by a further (marginally significant) finding; namely, that speakers' sensitivity to their partners' needs during planning depended upon the availability of feedback (Addressee-by-Novelty-by-Feedback Availability interaction: $\chi^2_{(1)} = 2.93, p = .087$). As indicated by Figure 3.5, speakers showed greater sensitivity to their partners needs when feedback was unavailable (simple interaction of Addressee-by-Novelty: $\chi^2_{(1)} = 7.43, p = .006$) than when feedback was available (simple interaction of Addressee-by-Novelty: $\chi^2_{(1)} = .15, p = .702$). However, even in the feedback-unavailable case, in which stronger partner effects were present, there was still no evidence that speakers scanned more when describing old referents to new (vs. old) addressees (1.1 vs. .8 gaze shifts, respectively; $\chi^2_{(1)} = 1.01, p = .315$). Instead, the interaction in the feedback-unavailable case seems to have been driven by speakers scanning the display more extensively when planning descriptions of new referents for the old ad-
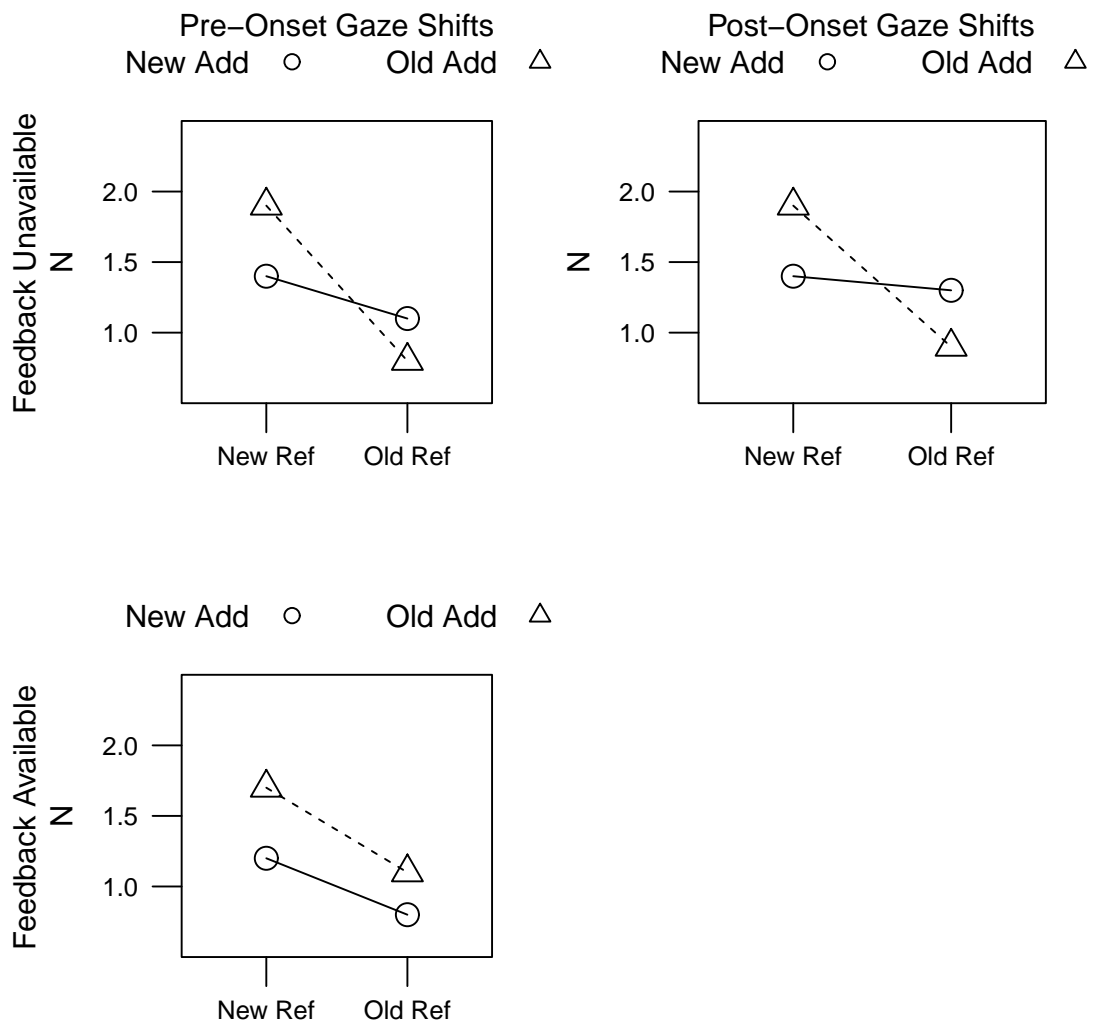
Figure 3.5: Number of pre- and post-onset gaze shifts for abstract trials.

dressee, although the effect was marginal (1.9 vs. 1.4 gaze shifts; $\chi^2_{(1)} = 2.81, p = .094$).

There was more reliable evidence for audience design effects in the rate of post-onset scanning, which reflects monitoring and adjustment. For this analysis, we considered only the feedback-unavailable condition, up to the point at which the speaker pressed the button. It is not sensible to analyze the data in the feedback-available condition, given the difficulty of distinguishing scanning related to considering display items versus scanning that reflects speakers following listeners' mouse movements. Speakers exhibited a stronger decline in scanning from new (vs. old) referents when they spoke to partners who shared their knowledge (about 50%; from 1.9 to .9) versus when they spoke to partners who did not (about 5%; from 1.4 to 1.3; Addressee-by-Novelty interaction: $\chi^2_{(1)} = 7.59, p = .006$). But again, the key prediction that they would scan more when describing old referents to new partners was not supported ($\chi^2_{(1)} = .18, p = .675$); nor was there evidence for a partner effect for new referents (1.9 vs. 1.4; $\chi^2_{(1)} = .84, p = .361$).

Thus far, all the evidence we have reviewed offers no evidence that speakers avoided underspecifying old referents through additional planning, which implies that these adaptations were accomplished through monitoring and adjustment. If this hypothesis is correct, we might be able to find positive evidence for this in a content analysis of the descriptions. Specifically, monitoring and adjustment predicts that a substantial number of descriptions of old referents to new addresses should be of the form "remembered expression (+ optional elaborated material)". We calculated the rate of such elaborated repetitions by comparing the description given for referents in the test phase to the terminal description given for that same referent in the training phase. This analysis revealed that elaborated repetitions were highest when speakers spoke to new addressees and feedback was available (Table 3.4). Between 28% (feedback unavailable) and 44% (feedback available) of all speakers descriptions to new addressees were

either repetitions or elaborated repetitions; this is compared to between 66% (feedback unavailable) and 73% (feedback available) for descriptions given to the same addressee.

In sum, when speakers talked about old referents, they successfully avoided underspecification, adapting to new addressees' need for additional information through monitoring and adjustment. When feedback was unavailable, they relied on self-generated assessments of these needs. When feedback was available, they minimized these self-generated assessments by relying on addressee feedback. There was no evidence that producing longer descriptions of an old referent for a new addressee required additional planning. Unexpectedly, audience design effects on planning were only observed for speakers descriptions of new rather than old referents: speakers gave longer descriptions of new referents for old (vs. new) addressees, taking about 400 ms longer to plan what they would say, and scanning the physical context more. These effects were not predicted, however, and so should be confirmed in later studies.

Finally, there was ample evidence that speakers took into account the interactional affordances of the situation, speaking longer and spending more time planning when feedback was unavailable. This is the first evidence we know of that interactional affordances of the conversational setting influence speech planning processes. However, there was little evidence that feedback interacted with audience design processes, apart from a marginally significant influence on pre-onset gaze shifts.

3.4   Discussion

This experiment considered in detail the use of over-specification and under-specification (through the use of concrete and abstract referents, and partner-switching), the opportunity for feedback, and eye-tracking data, in order to parse out the role of

attention and memory in speaker adaption and audience design. In addition, it offered evidence, through the manipulation of the availability of feedback, that the situational context can influence utterance planning.

Overall, as the experiment progressed, speakers tended to shorten their descriptions when describing abstract referents (c.f. Clark, 1996; Krauss & Weinheimer, 1964), consolidating their initially unwieldy descriptions over successive training trials. However, they could also still adapt when necessary and elaborate on their descriptions when using them for new addressees. While Clark (1992) demonstrated this, the results of the current experiment extend these findings by showing that speakers not only added new information incrementally in response to feedback (as was also primarily the case in the first experiment), but they also lengthened their descriptions in order to adapt to the unavailability of feedback. Lengthening, in this case, acted as insurance against the possibility that their utterance would be found to be inadequate. Over-specification in this case is a possible valid strategy because giving more information is potentially less costly in terms of listener comprehension than giving too little. Giving too little information may force the listener to resort to chance in order to pick out the target. On the other hand, while giving too much information may lead to an unintended implicature on the part of the listener since they are committing a violation of the Gricean maxim of quantity (Grice, 1975), this is less likely to reduce the chance of the target eventually being selected when the confusion is resolved. This may be why speakers tended to over-specify referents frequently, with under-specifications being comparatively rare.

Another explanation of why over-specification is more common, informed by the data, is based on the ability of speakers to adapt their utterances incrementally (Pechmann, 1989). It is interesting to note that speakers produced their lengthened descriptions of old referents to meet the needs of new addressees, without taking longer

71

to plan those utterances. This is because they could still rely on the same reduced utterance ("the crab") they had used at the end of training with their first partner to begin their description, and then elaborate on that description ("the crab, with claws on the sides") to make it more meaningful for the new, uninformed, addressee. This may be why elaborated repetitions (see Table 3.4) were more common for new addressees and relatively unnecessary for old ones. Under-specification is easily fixed through this elaboration, whereas as described above, over-specifications cannot be as easily corrected. Lastly, the eye-tracking evidence did not show an affect of the addressee manipulation prior to speech onset. This result, coupled with the similarity in planning time suggests that no additional audience design, driven by the partner manipulation, was happening during initial utterance planning.

The manipulation of feedback also speaks to the issue of when audience design is occurring. In the cases where speakers extended descriptions of old referents for the benefit of new speakers when there was no feedback, they did so as part of a self-prompted adaptation. The addressee could not inform them of the inadequacy of their utterance, and in this case adaptation is a clear example of audience design in action. When feedback is available, direct feedback from the listener allows for other-prompted adaptation, which by necessity comes after the initial articulatory process on the part of the speaker (see first and third position repairs in Schegloff et al., 1977). Speakers spent more time planning their descriptions when they knew they could not rely on feedback, but anomalously only for abstract referents. It may be the case that speakers found the chance for misunderstanding to be higher for the abstract items, and took their time in those trials relative to when they were referring to concrete items in order to reduce the possibility of miscommunication.

Viewing audience design as a form of expert performance offers a useful frame-

work within which we can begin to understand these results. Speakers, be it in the training phase of this experiment or in everyday discourse, can be seen as building up "routines" in memory that represent past useful solutions to problems such as the descriptions they have used previously to refer to a referent (Pickering & Garrod, 2005). Speakers build up a repertoire of such routines that can then be activated and deployed quickly during the process of language production, in order to quickly solve problems that are similar to ones that have been encountered recently. As a consequence of the local expertise developed over the course of a conversation, descriptions begin to take on properties more characteristic of proper names in discourse (Carroll, 1980). The process of reducing the length of an abstract referent's description into a compact form such as "the crab," is an example of this process in action. In this way, speakers avoid having to repeatedly recompute a description from scratch, which offers yet another avenue through which speakers can appear to have engaged in deliberative audience design, when in fact the relevant audience design processes had occurred much earlier in the discourse.

Chapter 4

Experiment III

4.1 Introduction

When deciding how to describe an object, a speaker first has to categorize that object in a way that is useful for them and hopefully useful for their partners. Sometimes, the taxonomic, established categorization is not well suited to the task and interlocutors may be forced to develop an ad hoc, or socially emergent category that is more suited to achieving their current conversational goals (Barr & Kronmüller, 2007; Barsalou, 1991). For example, when two people discuss where they wish to go out to eat they may form a category of acceptable restaurants on the spot that gets developed over the course of the conversation that is neither pre-existing in memory nor particularly enduring unless they frequently go out together. This socially emergent category then becomes the basis for future exchanges. What this means is that early on in a conversation, taxonomic categories are going to be activated more often and attract more attention. But as time goes by and speakers learn to put more emphasis on shared information, and other types of information are found to be in conflict with the goal of successfully referring to an object, the socially emergent categories should be used more prevalently. These socially

emergent categories would probably demonstrate some degree of being person-specific (Horton & Gerrig, 2005a), but that would not preclude their activation and propagation to new conversational partners if they were useful (Garrod & Doherty, 1994).

Reminiscent of the attention shifting hypothesis from the first experiment, people do learn to attend to features and objects that are relevant for completing their task, and ignore those that are not. Rehder and Hoffman (2005) conducted an experiment in which participants had to do a category learning task while having their eyes tracked. At the beginning of their experiment, participants would overtly attend to each of the features in the display and consider them all before making a decision. By the end of the experiment, participants learned to overtly attend only to those features that were necessary for the task of choosing which category was represented by the display. If one was relevant, they learned to attend just that one. If two were relevant, they learned to only two, and so on. It seems probable then that speakers will learn to attend primarily to those objects that are relevant to the task at hand. In a referential communication experiment involving perspective taking, as time passes and socially emergent categories become more dominant, attention should shift away from privileged competitors. By the end of a conversation they should be attended to about as little as an object unrelated to the target. This shift should also happen in situations in which the speaker does not initially have access to what the listener knows. As evidence about listener's knowledge, or lack of knowledge, becomes available they should gradually begin to focus on the stimuli that are most consistent with the listener's perspective.

The question then becomes, what factors influence the rate of learning socially emergent categories and the content of these ad hoc categories formed during conversation? One possibility is the potential for feedback. Closely related to feedback are the roles of the interlocutors: are they in a didactic situation in which the roles are fixed,

or is it a fully interactive dialogue in which there is turn taking? Unrestricted feedback can come in many forms. First, speakers can receive pragmatic feedback in the form of knowledge about the success or failure of the listener in establishing joint attention to a target object. It is implicitly accepted that pragmatic feedback is necessary for learning to happen, with positive feedback leading to a reinforcement of the speaker's current conceptualization, and negative feedback encouraging re-conceptualization. Second, the listener can give the speaker signals from which they can infer the listener's confidence about an interpretation. For example, Barr (2003) found that listeners will use speakers' confidence as a cue that the current referent is either a typical or atypical member of a category. In principle, speakers should also similarly use the listener's confidence in their selection of a target as a cue to how well their message was received.

Finally, listeners can give direct verbal feedback to the speaker. Schober and Clark (1989) conducted an experiment in which a participant acted as a non-interactive observer to a conversation. This participant attempted to do the same task the actual listener was doing, but without the ability to give feedback to the speaker. They found that the observers never aligned as closely to the speaker due to their inability to give feedback and have the speaker adjust to or correct their interpretations. The lack of verbal feedback can also make the speakers prone to give more information as a hedge against possible misunderstandings (Gann & Barr, 2012). For example, Krauss and Weinheimer (1966) found that when listeners could not give feedback, speakers tended not to simplify their utterances, but also kept them long in order to minimize giving too little information. While it is possible for alignment to happen without it, direct verbal feedback likely serves to speed up the process. Without it, speakers may not be able to easily settle on a shared perspective and fully acquire a socially emergent category that corresponds to the intersection of their knowledge with their partner's.

Role switching may serve a similar function. In a natural conversation, both participants are speakers and listeners. In this case, each are offering their own conceptualization of the referents they are bringing to each other's attention and are able to interactively sculpt each other's understanding. They are constantly giving each other positive evidence about what they know, and have full opportunity to give corrective feedback. In situations where one person is the primary speaker, there might be less incentive for them to align, and the speaker might rely on the listener to align to their knowledge.

In most referential communication experiments that manipulate common ground, knowledge is asymmetric in one direction: either the speaker knows more than the listener, or vice versa. This experiment focuses on a collaborative situation that is perhaps slightly closer to reality: a case in which the speaker and listener both have some knowledge that is privileged, and they must learn to account for this difference in knowledge. The major question is whether or not conversational partners in such a situation tend to form a socially emergent categorization of the stimuli that is consistent with a *union* of their perspectives, thus taking into account both their sets of privileged information, or an *intersection* of their perspectives, focusing only on that information that is mutually shared. If in a particular circumstance, speakers tend towards reconceptualizing the referent in terms of an intersection between their knowledge and their partner's, will speakers then begin to ignore competitors that are not relevant to that overlap? Additionally, is their tendency to adopt one scheme or another influenced by the form of feedback they experience: either explicit feedback, or implicit feedback through role switching? Direct feedback should allow addressees to clearly indicate to the speaker what their informational needs are, and should thus be associated with rapid alignment between the interlocutors. Learning about a partner's perspective in the role switch-

ing condition presumably relies on more indirect learning, and may not have the same impact as direct feedback.

4.2 Method

4.2.1 Participants

The participants were drawn from a pool of undergraduates at the University of California, Riverside. A total of 32 students participated in the study, forming 16 dyads. The participants were evenly divided between males and females, who were also assigned evenly across the between-subjects conditions. For half of the dyads, the participants were assigned to either the role of the director or the matcher. The other half of the dyads did not assign the participants to fixed roles, but rather they switched roles after each trial of the experiment.

4.2.2 Apparatus

The apparatus is the same as it was in the previous experiments.

4.2.3 Materials

The experimental stimuli consisted of 32 sets of five items each. Each stimulus item was a 300 x 300 pixel colored bitmap of an object placed on a black background (for a list, consult Appendix C). The sets were constructed such that there was a target object, three competitor objects that are typically referred to by the same name, and an unrelated filler item. The competitor objects contrasted with the target object along the dimensions of size, openness, and material (or color). The target was always consistently open and larger; each of those two dimensions were visible to only one of the

Figure 4.1: Each of the displays above corresponds to the perspective of one participant in a single trial. In this example, the target would be the closed metal trashcan in the upper left corners. The display on the left contains a competitor based on openness: an open version of the target. The display on the right shows a competitor based primarily on material: the plastic trashcan. Each display has a competitor in common based on size.

participants. Material differed from object to object; this dimension was always shared between the participants. So for example, one participant may have seen a target that was a large open trashcan, a smaller identical competitor, a plastic trashcan, and an unrelated fourth object. The other participant would see the same target, the plastic trashcan, the same unrelated competitor, but would have a closed competitor that is otherwise as physically similar to the target as possible. See Figure 4.1 for an example of both the participants' perspectives and the dimensions used.

4.2.4 Procedure

Participants were assigned to either the role of director or matcher by having one of the participants choose a face down card at random that specified a role. If the dyad was participating in the condition in which their roles would switch throughout the experiment, it was explained to them that they were picking an initial role. For

79

the switching participants, the participant who was eye-tracked was always the initial director (the consequences of this for the analysis are discussed in 4.2.6 and 4.3).

The experiment was described to the participants as a simple communication game that they would be playing together in which the director would be describing a target object such that the matcher could pick out which object on their screen. They were shown a set of example stimuli demonstrating the perspective of both partners. The sample stimuli did not include competitors that varied along the privileged dimensions, only one that contrasted along the shared dimension. The participants were instructed that the director's view would include an object that is highlighted in red and that this would signal to them that this is the target object to be referred to. The matcher would listen to the director's description and select the object they think is being talked about on their screen by clicking on it with their mouse.

4.2.5 Design

The experimental variables manipulated in the design of the experiment included whether the partners switched roles or not, and level of feedback (2x2 between-subjects design). The dependent variables are accuracy (whether the addressee chooses the right target), speech onset time, use of an adjective corresponding to the speaker's privileged perspective/dimension, use of an adjective corresponding to the addressee's privileged perspective, use of an adjective corresponding to the shared perspective, and the proportion of time the speaker spent gazing at the shared and privileged competitors.

4.2.6 Analysis

The director's speech was recorded, transcribed, and coded for accuracy, speech onset, and the use of modifiers matching each of the three dimensions on which the

competitors contrasted with the target. The eye-tracking data were coded as number of frames (sampled at 60 hz) that the director spent looking at each object prior to speech onset. Eye-tracking data for trials in which the participants switched roles was removed from the data set due to the inability to make an apples to apples comparison across the condition due to the current director only contributing eye-data for half the trials in the switching condition. Eye-tracking data in half of those trials was collected on the listener due to the partner-switch manipulation, and while some interesting questions could be asked about the gaze pattern of the listener, there were too little data for a reliable analysis.

The analyses were conducted using the same modeling techniques described in the analysis of the first experiment.

4.3 Results

Thirteen of 512 (2.5%) trials were dropped from the analyses due to experimenter and procedural errors. Due to the absence of eye-tracking data for the second participant in the switching group, and the inclusion of trial order as a factor in the analyses, only the first sixteen trials of the non-switching group were analyzed and compared with the performance of the eye-tracked partner in the switching group. This adjustment further reduced the number of observations per experimental session to 16, for a remaining total of 363 observations. Preliminary analyses suggested the possibility of an order effect depending on which partner was the director first in the switching group, which also lead to the decision to remove the data for participants who weren't eye-tracked even in the speech-only analyses in the interest of equivalence between the conditions.

Table 4.1: Summary table for accuracy.

Factor	Estimate	$\chi^2_{(1)}$	p
Feedback (F)	2.295	26.934	$< 0.01^{**}$
Switch (S)	-0.236	0.389	0.533
Order (O)	0.127	11.990	$< 0.01^{**}$
S:F	-0.821	1.224	0.269
S:O	-0.156	4.243	0.039^*
F:O	0.086	1.220	0.269
S:F:O	-0.289	3.600	$0.058^{\dagger}$

$^{\dagger}p < 0.10$; $^*p < .05$; $^{**}p < .01$.
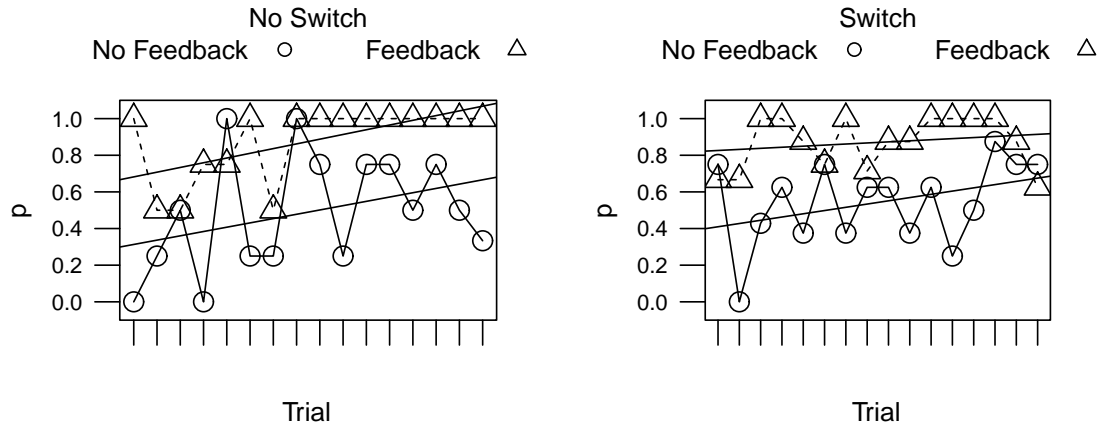
Figure 4.2: Proportion of trials in which addressees chose the correct target.

Overall, participants were fairly accurate, with the listener correctly identifying the referent in 72.55% of trials. Switching $(\chi^2_{(1)} = 26.934, p < 0.01)$ influenced the accuracy of pairs, with role switching being associated with slightly poorer performance, probably due to slightly increased task difficulty associated with exchanging roles between trials (see Figure 4.2). This effect of role switching interacted with trial order, such that the effect of role switching's impact grew over the course of the experiment with non-switchers hitting ceiling performance relatively quickly $(\chi^2_{(1)} = 4.243, p < 0.05)$.

4.3.1 Speech Analysis

Potentially, speakers could refer to three dimensions in their descriptions of the referent corresponding to the contrasts between the referent and the shared competitor, the speaker's privileged competitor, and the listener's privileged competitor. At the beginning of the experiment speakers are much more likely to use the first two by virtue of their lack of access to the third, but as the experiment progresses and evidence for that dimension becomes available it should be seen to rise. The use of adjectives/modifiers associated with these dimensions speaks to the underlying scheme the speaker has about what information is necessary to identify the target. Because the common competitor is always necessary (and indeed, is referred to in 88% of trials) its use is a useful check on whether a speaker understands and is faithfully attempting to do the task, though less interesting from a more in depth analytic point of view. Since the underlying hypotheses are focused on how speakers reconcile their privileged knowledge with that of their listeners, how they used modifiers related to the two privileged contrasts is more theoretically interesting.

Use of a modifier relating to the speaker's privileged competitor was influenced by whether or not the participants were switching roles ($\chi^2_{(1)} = 3.977, p < 0.05$). This effect interacted with trial order ($\chi^2_{(1)} = 3.798, p = 0.051$), such that the size of the switch effect grew over the course of the experiment (see Figure 4.3). For the non-switchers, use of the speaker's privileged competitor stayed relatively close to ceiling over the course of the experiment whereas its use dropped over time for partners who switched roles. This may be because, for a speaker in the no-switch condition, their privileged competitor was always salient to them, whereas in the switching condition speakers experienced trials in which its presence was not linked to production. Since the

83

Table 4.2: Summary table for speaker's privileged modifier use.

Factor	Estimate	$\chi^2_{(1)}$	p
Feedback (F)	-0.186	0.157	0.692
Switch (S)	-0.978	3.977	0.046*
Order (O)	-0.004	0.014	0.905
S:F	1.110	1.365	0.243
S:O	-0.130	3.798	$0.051^{\dagger}$
F:O	0.010	2.293	0.130
S:F:O	-0.219	2.771	$0.096^{\dagger}$

$^{\dagger}p < 0.10$; $^{*}p < .05$; $^{**}p < .01$.

Figure 4.3: Proportion of speaker utterances that incorporate information about the dimension that is privileged from the speaker's perspective.

use of the privileged competitor was not strongly associated with accuracy ($\bar{r}_{\phi} = -0.028$, p = 0.595), there was not necessarily a strong reason to reduce its salience to speakers in the non- switching condition in the absence of another role.

On the other hand, the speaker had more substantial motivation to learn and use the listener's privileged competitor since its use was strongly associated with success in the task ($r_{\phi} = 0.310, p < 0.01$). In their use of modifiers that matched the listener's privileged competitor, speakers were significantly influenced by an interaction between the feedback and switching conditions, such that the effect of feedback was larger when

Table 4.3: Summary table for listener's privileged modifier use.

Factor	Estimate	$\chi^2_{(1)}$	p
Feedback (F)	4.650	24.278	$< 0.01^{**}$
Switch (S)	0.484	0.457	0.499
Order (O)	0.264	24.582	$< 0.01^{**}$
S:F	-3.097	5.701	0.017^*
S:O	-0.219	3.324	$0.068^\dagger$
F:O	-0.032	0.055	0.815
S:F:O	0.148	0.317	0.573

$^\dagger p < 0.10$; $^* p < .05$; $^{**} p < .01$.

there was no role switching ($\chi^2_{(1)} = 5.701, p = 0.17$). It is interesting that there doesn't appear to be an additive effect between the two routes to getting partner feedback associated with the two factors, with the greatest degree of use of the listener's privileged competitor being when there was no role switching (see Figure 4.4). It's possible that the switching condition in this case reduced the pressure on the speaker to unilaterally integrate the perspective of their partners that is granted primarily through the channel of direct feedback.

Indirect feedback about their partner's knowledge may not figure into their subsequent production in the switching conditions. This is because some of the information is coming during trials in which they do not have to integrate this information into an utterance themselves, and for which it is unnecessary from their perspective, thus reducing its salience.

While the focus in this experiment is on the effect of feedback as a channel from which speakers will get information about the divergent perspectives of their partners, speech onset offers a window into feedback's anticipatory effect in regards to how much planning speakers have to do when they are aware of the possibility that their addressee can make a clarification request. Much as in the experiment from the previous chapter,

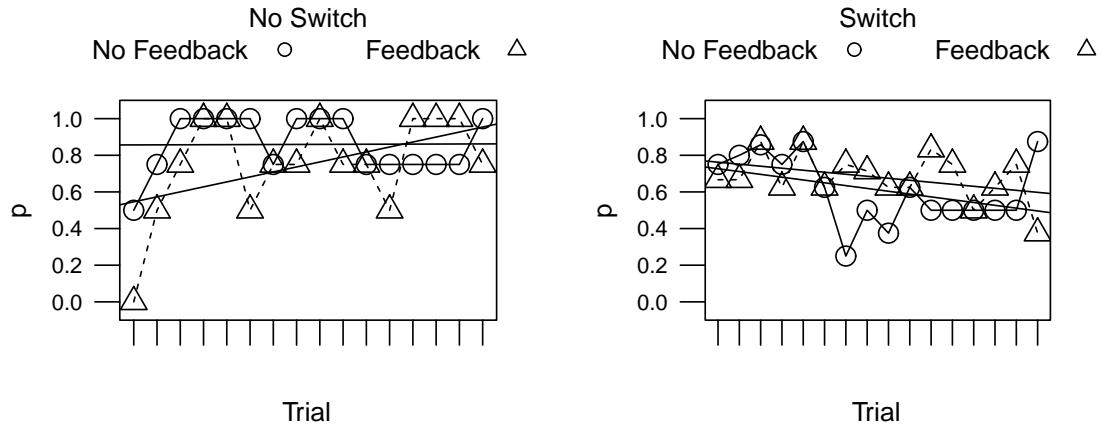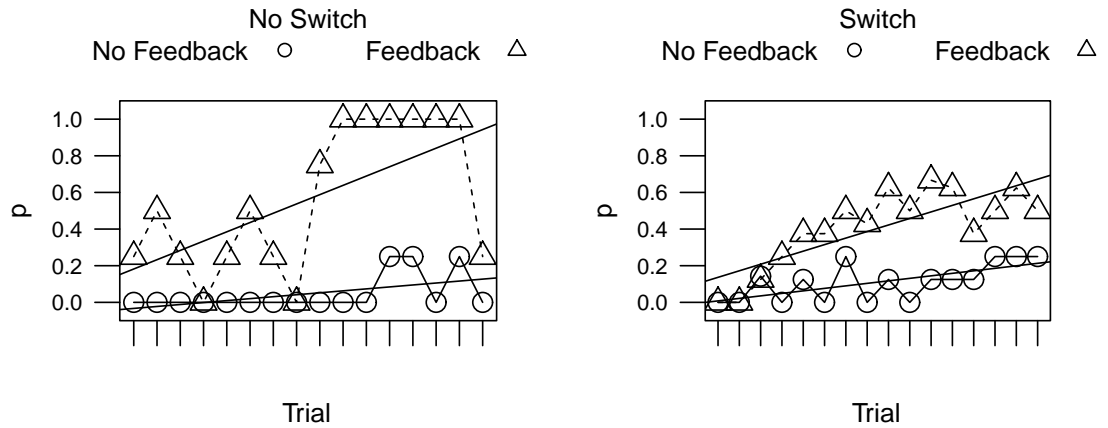Figure 4.4: Proportion of speaker utterances that incorporate information about the dimension that is privileged from the addressee's perspective.

Table 4.4: Summary table for speech onset.

Factor	Estimate	$\chi^2_{(1)}$	p
Feedback (F)	-0.487	11.163	$< 0.01^{**}$
Switch (S)	0.061	0.004	0.950
Order (O)	-0.020	24.195	$< 0.01^{**}$
S:F	0.288	1.145	0.285
S:O	-0.001	0.039	0.844
F:O	-0.009	1.283	0.257
S:F:O	0.025	2.492	0.114

$^{\dagger}p < 0.10;\ ^{*}p < .05;\ ^{**}p < .01.$

planning (time to onset of speech) was influenced by the opportunity for feedback ($\chi^2_{(1)} = 11.163, p < 0.01$) with planning time decreasing with the presence of feedback. Similar to previous findings, this is likely due to reduced pressure on the speaker to carefully craft their utterances due to the lack of chance for correction. Time needed for planning showed a significant decreasing trend over the course of an experimental session ($\chi^2_{(1)} = 24.195, p < 0.01$), which is consistent with a general increase in familiarity with the task at hand.
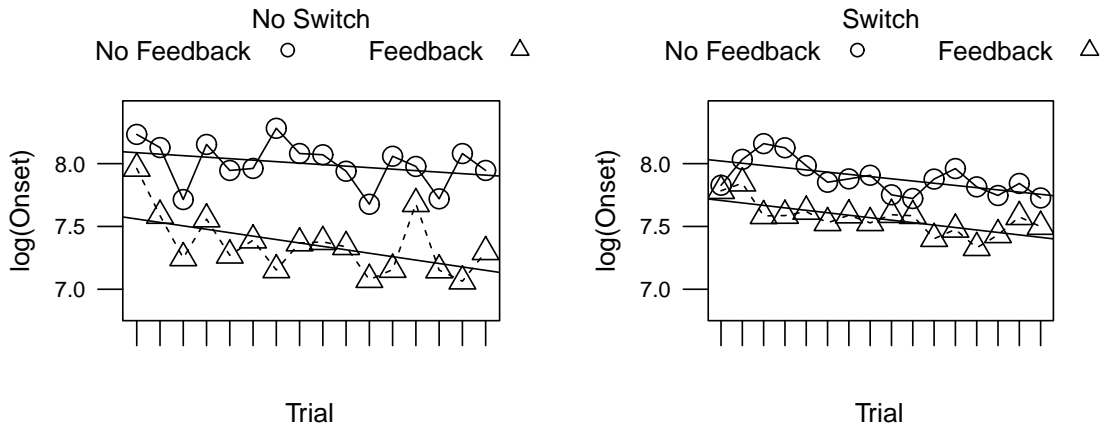
Figure 4.5: Speech onset time plotted as a logarithmic scale.

Table 4.5: Summary table for gaze to the privileged competitor.

Factor	Estimate	$\chi^2_{(1)}$	p
Feedback (F)	0.318	4.820	0.028*
Switch (S)	0.046	0	1
Order (O)	-0.004	0.341	0.559
S:F	-0.065	0	1
S:O	0.002	0.209	0.885
F:O	0.011	0.451	0.502
S:F:O	0.027	0.687	0.407

$^\dagger p < 0.10$; $^* p < .05$; $^{**} p < .01$.

4.3.2 Gaze Analysis

One of the primary claims of the hypotheses motivating this experiment is that patterns of gaze during planning may reflect the underlying categorical structure a speaker is considering. Rehder and Hoffman (2005) found that in category discrimination tasks participants learned to only attend to the features in a display that are required for discriminating category membership for the pattern as a whole. Likewise, it was hypothesized that gaze in a referential communication task would demonstrate the

Figure 4.6: Proportion of gaze time directed at the privileged competitor prior to speech onset, transformed into empirical logits. For reference, 0 is equivalent to 50% and -4 is approximately equal to 1% of the time.
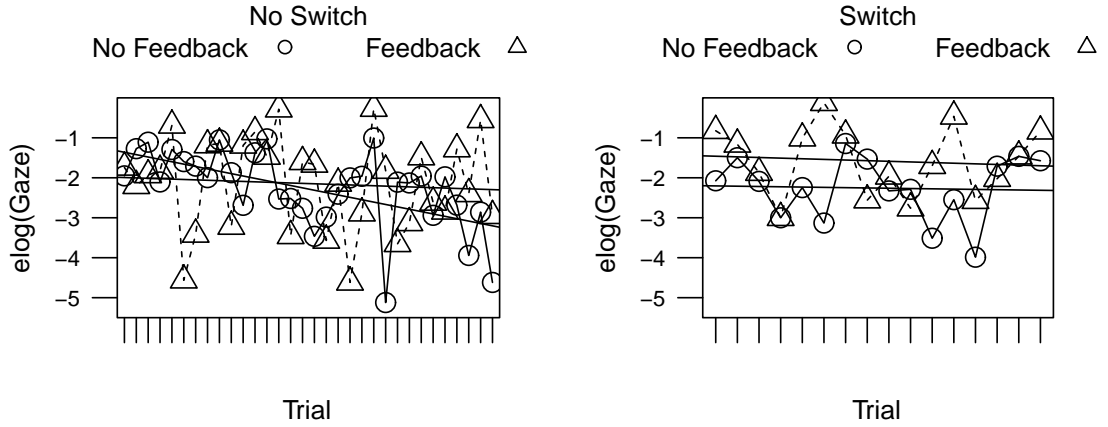
Table 4.6: Summary table for gaze to the shared competitor.

Factor	Estimate	$\chi^2_{(1)}$	p
Feedback (F)	0.437	8.651	$< 0.01^{**}$
Switch (S)	-0.303	4.339	0.037
Order (O)	0.001	0.020	0.887
S:F	-0.342	1.578	0.209
S:O	0.011	0.321	0.571
F:O	-0.047	4.065	0.044^{*}
S:F:O	-0.010	4.777	0.029^{*}

$^{\dagger}p < 0.10$; $^{*}p < .05$; $^{**}p < .01$.

shifting category membership status of the objects under consideration by the speaker.

Attention directed to the privileged competitor (measured as the proportion of pre-onset fixation time) doesn't appear to be influenced by the same factors that influence the use of the privileged competitor. In this case, the only factor of influence is feedback ($\chi^2_{(1)} = 4.820, p < 0.05$), with feedback being associated with increased attention being directed to the privileged competitor. Theoretically, attention to the shared competitor should be relatively insensitive to condition due to its constant use
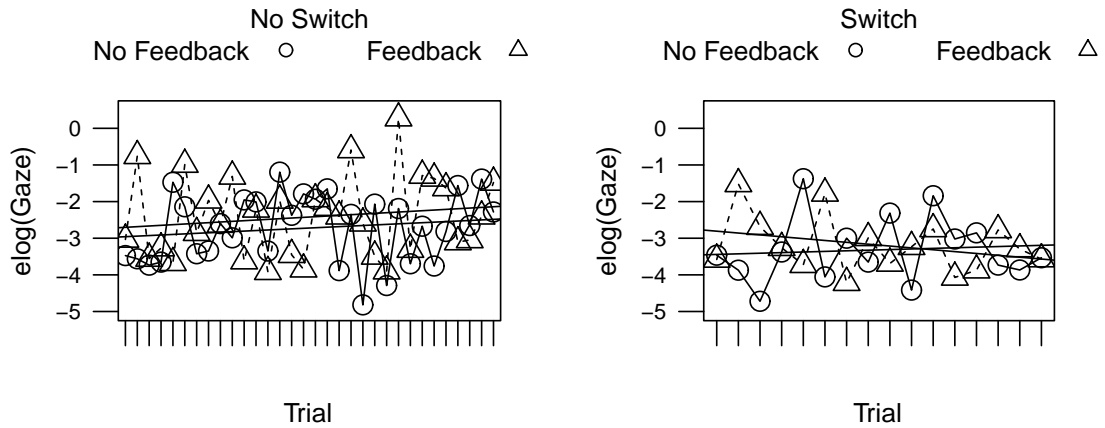
Figure 4.7: Proportion of gaze time directed at the shared competitor prior to speech onset, transformed into empirical logits. For reference, 0 is equivalent to 50% and -4 is approximately equal to 1% of the time.

regardless of circumstance, but it appears to be more heavily influenced by experimental factors than attention to the privileged contrast. This is demonstrated by a three-way interaction between switching, feedback, and trial order ($\chi^2_{(1)} = 4.777, p < 0.05$), which appears to show an increase in the effect of feedback as trial order increases, but only in the role switching condition. This suggests that addressee feedback caused speakers to focus more on the shared competitor over time, but it is not clear why this would be the case.

4.4 Discussion

Speakers and their addressees tend to align their conceptual and semantic representations over the course of conversation, however the final form that these representations take may depend on the characteristics and constraints under which the interaction is undertaken. In this experiment, each participant had information that was privileged and the goal was to determine what effect conversational constraints,

such as role constancy and the opportunity (and actuality) of feedback, had on the nature of the interlocutors' apparent representations as evidenced by the form of their utterances. These changed representations are supposed to reflect the creation of socially emergent categories that arise out of the alignment process.

Interestingly, independent of all of the experimental conditions, the most successful descriptions (in regards to whether the listener was able to correctly identify the referent) included the listener's privileged competitor. This is why performance was lowest by far in the most feedback-impoverished condition: when participants were not allowed feedback, nor did they switch roles. This is notable because the task was, in principle, solvable if the addressee aligned to the speaker and took the speaker's ignorance of their privileged competitor into account. Just as the speaker had the opportunity to learn what the listener's perspective was through feedback, the listener could infer the speaker's perspective through their production (or in the case of role switching, when they were in the opposite role). A speaker's use of their own privileged perspective had relatively little bearing on the outcome of an individual trial.

Speakers seemed to acquire information about the addressee's perspective most effectively through direct, corrective feedback (as in this case the opportunity for feedback has to be actualized for such learning to happen). Despite the opportunity to model their partner's knowledge when they occupied the directing role in the role switching condition, speakers appeared to primarily acquire this information through direct feedback such as the listener asking about their privileged dimension. However, the significant role switching by feedback interaction speaks to a boosted effect of feedback when partners did not switch roles. This is interesting because it suggests that perhaps it is more beneficial to have role constancy and consistent feedback than to experience both roles and be in their partner's shoes. On the other hand, a speaker's use of their own privileged

90

competitor seemed to be sensitive only to role switching, with use staying at a constant, high rate when partners have stable roles, but declined steadily when they switch roles. One possible explanation for these results is that addressee feedback is more likely to be about their communicating the dimension that is salient to them, and not referring to the one that is not. As discussed in the last chapter, under-specification is potentially more damaging (but amenable to feedback correction), than over-specification. So direct feedback would put an upward pressure on the use of the addressee's perspective, but be neutral in regards to the speaker's.

Thus when speakers are role switching and have feedback, they are likely to align in a way that favors the addressee through the speaker's increasing use of the listener's privileged dimension. This suggests these participants are moving toward a conceptualization of the shared perspective that is a union of their perspectives: the speaker makes use of adjectives related to both of partners' privileged dimensions. As seen in previous experiments such as Wardlow Lane et al. (2006), anything that raises the salience of a particular competitor is likely to increase mention of that competitor, which might explain why the privileged competitor is slow to become extinct. Additionally, because its use doesn't seem to have a negative effect, there is little pressure to reduce its use absent another factor. Participants in the other conditions more heavily favor a "union" of their perspectives, in which they combine the use of both perspectives.

The online measures paint a more mixed picture. The effect of speech onset in regards to feedback seemed to mirror that which was seen in the prior experiment. When speakers were not permitted the opportunity for direct advance feedback, they took a little longer to plan their utterances. The eye-tracking data were less clear. At the outset it was anticipated that the attention paid to the competitors would help reveal the underlying conceptual structure the speaker was consulting in order to formulate

91

her utterances (Rehder & Hoffman, 2005). However, since the listener's competitor is never visible to the speaker, there was no way to match up eye-movements with its consideration. In addition, the relative salience of the competitors on the screen was under different selective pressures than the features in the categorization task of Rehder and Hoffman (2005), as additional decisions had to be made about how to describe the contrast rather than making a simple judgment of category membership. Attention to the shared competitor, for example, should be relatively equal across conditions, but it actually showed a three-way interaction in which a significant feedback effect emerged over time, but only in the switching condition. It's unclear what exactly this would mean, since use of the shared competitor adjective was universally high (and universally required to distinguish the target). On the other hand, despite the fact that privileged modifier usage was most sensitive to feedback, gaze toward the object itself seemed to be governed by whether or not the dyad role switched.

Chapter 5

General Discussion

This dissertation began with an overview of several approaches to explaining the phenomena of audience design: Initial Design, Monitoring and Adjustment, the Ordinary Memory hypothesis, and aspects of Interactive Alignment theory. Rather than being competing complete theories of audience design, each has something to offer to a hybrid approach that recognizes that audience design is not a single process, but rather emerges as the result of several mechanisms that work together to allow a speaker to incorporate her audience's perspective into her descriptions of referents.

The first experiment sought to describe an additional mechanism, the Attention-shift Hypothesis, through which speakers might be able to preemptively direct attention away from privileged competitors in order for that information to not influence, and thus be leaked by, speech production. This experiment attempted to focus on the role of attention in helping select objects to be considered as input to further audience design. By diverting attention away from where privileged information is expected to be, speakers can show what appear to be early effects of audience design in their utterances even when they are not actively engaging in audience design. The experiment

also demonstrates that when the common ground status of the objects changes unexpectedly, speakers are unable to immediately adjust to the new perspective. This argues against an automatic and early effect of audience design. The manipulation of speaker attention led to a reduction in object salience, demonstrating a complementary effect to the ironic process seen in Wardlow Lane et al. (2006).

The second experiment investigated the role of the availability of feedback in message formulation, and used concrete and abstract stimuli to elicit over- and under-specification in dialogue in order to demonstrate that speakers rely on memory routines when designing utterances. The most important contribution of this line of work is to recast audience design as a limited form of expert behavior. Speakers, over the course of a conversation, build up a repertoire of expressions that have been successful previously in the discourse and are thus likely to be useful again. This leads to a bias towards over-specification in referential descriptions because of the ease with which speakers can add to a routine, as opposed to completely redesigning it or attempting to undo what was said. Backing up some of the claims about attention made in the first experiment, this experiment found that there was no audience design driven by the partner manipulation (in which the partner was switched between the training and test phases) in early utterance planning. Self-initiated adaptation, or audience design, in the circumstances described in this experiment are reserved for limited circumstances in which a speaker has no opportunity for feedback from their addressee. Importantly, the adaptation in this experiment provided evidence that the planning process is influenced by the interactional affordances of the situation.

The final experiment was concerned with investigating the nature of the shared representations, or socially emergent categories, that interlocutors build (c.f. Barr & Kronmüller, 2007; Pickering & Garrod, 2004) and align to in the course of a novel

referential communication task. It sets up a situation in which both partners in the communication game have privileged information. Here success was contingent with at least one of the interlocutors adjusting their conceptualization of the referent in response to information learned over the course of the experiment and not just a single trial. The manipulation of participant role (director vs addressee) and the potential for feedback both established different ways in which interlocutors could share their privileged information. It was shown that this transfer of information was best achieved through direct feedback. However, it was also seen that participants who switched roles saw an enhanced effect. It was found that generally the most successful strategy was to adjust to the knowledge of the addressee, although consequences for the speaker's use of their own privileged information were more mixed.

Each of the experiments was designed not only to shed light on a mechanism through which audience design can be achieved, but also to ground the process of audience design firmly into a more general cognitive framework. One such candidate framework is the Instance Theory of Attention and Memory (ITAM; Logan, 2002), representative of a large family of models, that attempt to intimately tie together attention, memory, and categorization into one broad theory: the same mechanisms at the heart of each of the experiments above.

5.1 Grounding Audience Design in a Broader Cognitive Theory

The primary task that ITAM is concerned with is selecting objects to attend to and categorization, which are integral parts of the enterprise of producing utterances in a referential communication paradigm. It must be made clear that ITAM is pri-

marily a mathematical and conceptual model suited to a different primary task, and the goal at present is to fit audience design into its conceptual framework, or at least borrow elements that may be adapted for use in understanding speech production and utterance planning. The project of integrating the two is very important because, as mentioned before, conversation is one of the primary avenues through which we not only deploy our accumulated conceptual knowledge, but through which we adjust and add to that knowledge (Barr & Kronmüller, 2007). Selecting and attending to objects occurs in parallel with utterance planning and formulation. The processes modeled by ITAM can be seen as part of what feeds input into the initial stages of language production, and how decisions about which conceptualizations get used are made. Logan (2002) proposed that, "To attend is to categorize; to categorize is to remember; to remember is to attend," suggesting that as a speaker attends to a target object, she is also simultaneously categorizing it. In previous work using eye-tracking, gaze was correlated with the retrieval of a name for an object (Griffin, 2004). Inspired by Rehder and Hoffman (2005), Experiment III sought to observe a relationship between the use of adjectives associated with particular dimensions and speakers' overt attention to the competitors that contrasted with the target along those dimensions. Predictions like this are part of what models like ITAM seek to describe. Indeed, Rehder and Hoffman (2005) also look to ITAM for explanations of their results, among similar models. I focus on ITAM here, because the cognitive phenomena it attempts to integrate–attention, categorization, memory–are integral to the major themes of each of the three experiments presented in this dissertation.

Selecting an object, and retrieving a concept associated with it, are fundamentally decision processes. As discussed in regards to Experiment II, one way of viewing the decision making process about which conceptualization to use in production is as a

race model. Automatic memory retrieval and inference processes compete with slower cognitive processes such as a deliberate consideration of common ground. The idea of viewing audience design as a form of expert behavior is based on the assumption that accumulated expertise will further rig the race in favor of descriptions made highly available through experience. ITAM's preferred decision rule is to use a random-walk model, based on the exemplar-based classification model of (Nosofsky & Palmeri, 1997) but which has been applied to other forms of decision making. This conception of a random-walk model subsumes the concept of the race model while extending it to broader circumstances. A random-walk model of decision making suggests that the race runs until a solution has outpaced its competitors by a particular degree of confidence. When the degree of confidence required is low, a random-walk model acts like a race model: the most swiftly produced solution wins. This decision criterion thus corresponds to the willingness of the agent to trade speed for accuracy. Thus, the degree of confidence required is dependent on situational factors; one such factor relevant to the work presented here is the potential for feedback. One explanation for some of the effects of the potential for feedback on utterances seen in Experiments II and III, is that the expectation of feedback changes the confidence required for a solution to be settled on. When feedback is plentiful and expected, speakers can lower the confidence required to make a decision and speech onset time decreases on average. In these cases, they rely more on highly available conceptualizations of the referents left over from previous instances. When feedback is curtailed, speakers act more conservatively, taking longer to produce their utterances and are more likely to reconceptualize them in order to meet the needs of their audience. Thus, when there is little cognitive load, and uncertainty is high, speakers may be observed to act in a manner more superficially consistent with the Initial Design hypothesis. However, when speakers are more confident, practiced, or

under pressure to perform they will show more evidence on relying on swiftly available memory routines.

It is also possible that other factors, such as task uncertainty and unfamiliarity may also effect speaker confidence. ITAM includes many parameters that can be thought of as having great relevance to audience design, but for which more direct experimental work would be needed to define their impact in a referential communication context. Three parameters that are important within ITAM that could have impact on selection and ultimately speech, are: the evidence that an object is a member of a particular category, the importance of choosing objects that are a member of a particular category, and the priority of a particular property in making that selection (described within Logan (2002) as η, β, and π, respectively). Each acts as a measure of bias toward making a particular selection, and are relevant to different properties of the environment and speaker at hand. Evidence that an object fits a particular category is drawn from perceptual information, and prior history: greater expertise in making that discrimination will increase the chances of making the appropriate selection. The importance of choosing an object that is a member of a certain category is controlled by the demands of the task, and again by prior experience of the speaker. Situational factors, such as the presence or absence of feedback, as we have seen before, could also be seen to have an influence. The last parameter, which sets a priority on making a selection based in part on a particular property is relevant because it offers an explanation for why speakers may include information, such as color (Mangold & Pobel, 1988), which may not be strictly needed for disambiguation of a referent. Speakers may include color to help the listener more quickly narrow down the alternatives they are faced with so that they can more quickly identify the referent. This is also an example of how these parameters aren't restricted to a consideration of the speaker: conversational partners sculpt each

other's expectations and biases throughout the course of a conversation. A potentially exciting future area of interest may be to not only formalize a model of audience design, but also to instantiate it within a multi-agent context.

5.2 Future Directions

There are two theoretical gaps to fill in the current understanding of the theories considered thus far. One concerns how hypothetical changes in underlying conceptual structure impact the surface structure of the utterances through the message they intend to convey. The second is the automatic or controlled nature of the acquisition of implicit ground, and the implications that has for the role of implicit and explicit learning in audience design and conversational interaction. Following up on these two lines of research would also help specify areas with which audience design can be linked to broader theories of cognition such as ITAM.

The use of different types of modifiers was touched upon briefly in the first experiment when a distinction was made between pre-nominal modification (e.g. "the **blue** giraffe") and post-nominal modification (e.g. "the giraffe, that is **blue**"). In that case, the data were clear that the instances of post-nominal modification were mostly explainable as late repairs driven by partner feedback through hesitation due to the situation and the eye-tracking data (recall Brown-Schmidt & Tanenhaus, 2006; Brown-Schmidt & Konopka, 2008). However, there are other proposals about reasons [English] speakers might make a distinction. Saylor (2000) suggests that the difference is that pre-nominal modifiers suggest that the feature is relatively time-stable and enduring, whereas post-nominal modifiers are less time-stable features for the category being identified. Markman and Makin (1998) hypothesize that pre-nominal modifiers typically

refer to alignable differences between objects in a category, that is differences among features that they share. Post-nominal modifiers refer to non-alignable differences; they refer to parts of one object that are not typically features of another object of the same category it is being compared to. Another way of looking at the difference might be to focus on the salience of the feature being referred to by the adjective. Highly salient features might be referred to early, as pre-modifiers, whereas less salient features are referred to later. This explanation fits in well with the observation that superlative adjectives, such as smallest or loudest, are typically placed in a pre-nominal position (Smith, 1961).

Each of these explanations presupposes a difference in the underlying conceptual structure being relayed, with the difference between them concerning what exactly the nature of the difference is. It is possible that upwardly or downwardly manipulating the relative salience of these dimensions in memory may be reflected in the structure of referential descriptions. Changes in underlying concepts being employed in speech should be reflected in how modifiers are ordered within utterances. Post-nominal modification may be a side effect of late attention to the competitor (first experiment), or incremental modification (second experiment), but in the same vein it should, in theory, be possible to shift the usage of those modifies back in the direction of pre-nominal modification by increasing its salience, much as how utterances were found to be shorten and streamlined over time for the abstract items in the second experiment. Intimately tying utterance form to underlying conceptual structure would help contribute to an understanding of how categorization affects speech, and in turn influences audience design by altering the message that is being conveyed. Understanding how differential modification impacts the message being conveyed would also have consequences for understanding how over- and under-specification impact addressee comprehension.

Each of the experiments in this dissertation have been primarily focused on how speakers can appear to be engaging in audience design in the moment, but are actually engaging more heuristic style processes that capitalize on prior experience. This is similar in some ways to the discussion of implicit common ground as defined by Pickering and Garrod (2004), although they make the claim that implicit common ground is built up automatically and resource-free over the course of dialogue. It's interesting to note that the role switching condition in the final experiment of this dissertation appeared to offer evidence that this might not necessarily be the case. When it was a partner's turn as speaker, it was expected that they would have had some benefit of having previously listened to their partner's use of their privileged dimension, even if only through priming. However, switching seemed to not have a positive effect on learning a partner's privileged dimension.

There are some more theoretical reasons to doubt that implicit common ground is built purely automatically, following evidence cited by Fincher-Kiefer (2001) that situation model construction and updating relies on some degree of visuospatial processing and is vulnerable to a visuospatial load. While we have attempted to focus on a more instance-based instantiation of memory, it's still possible that load can interfere with these similar processes. In a response to their article, Pickering and Garrod (2004) comment on the possibility that two mechanisms could be involved: priming as conceived of as spreading of activation, or priming as implicit learning. The implicit learning account of priming in language (c.f. Bock & Griffin, 2000; Chang, Dell, Bock, & Griffin, 2000; Chang, Dell, & Bock, 2006) might be more compatible with the attention shift/category learning hypotheses stated previously, especially since implicit learning also shows some vulnerability to load (Cleeremans, Destrebecqz, & Boyer, 1998), much like situation model updating. The approaches, as has been assumed above, may be complementary.

It would be interesting to conduct research in order to distinguish them more cleanly, by manipulating cognitive load. In this case, however, it may be best not to use time pressure as in Horton and Keysar (1996), but rather something similar to the interleaved dual-task as seen in Roßnagel (2000).

Appendix A

Experiment I Stimulus List

Table A.1: All targets and competitors were identical except the image varied in size.

Target/Competitor	
Spoon	Fan
Pine cone	Key
Scissors	Pumpkin
Umbrella	Airplane
Book	Vase
Car	Leaf
Butterfly	Anchor
Gun	Binoculars
Carrot	Camel
Candle	Funnel
Cat	Gloves
Toaster	Bike helmet
Trash can	Koala
Pillow	Peanut
Television	Desk lamp
Camera	Salad tongs
Pencil	Spindle
Clock	Thermometer

Appendix B

Experiment II Stimulus List

Table B.1:

Target	Competitor	Modal response
Candle	(Not melted) candle	Unmelted candle
Key	(Old-fashioned) key	Modern/new/gold key
Knife	(Swiss Army) knife	Knife with the brown handle
Trash can	(Metal) trash can	Plastic/white trash can
Spoon	(Large slotted) spoon	Small spoon
Guitar	(Electric) guitar	Acoustic guitar
Carrot	(Cartoon) carrot	Little/real carrot
Gorilla	(Young/brown) gorilla	Black gorilla
Gun	(Toy) gun	Real gun
Leaf	(Dark green three pointed) leaf	Dark green leaf
Rose	(Wilted; stem not visible) rose	Rose with a stem
Marker	Marker (with cap off)	Marker with a cap
Screwdriver	Screwdriver (with black on handle)	Screwdriver, all red handle
Backpack	(Blue) backpack	Purple backpack
Clamp	(Open) clamp	Closed clamp

Appendix C

Experiment III Stimulus List

Table C.1: Each of the targets had three competitors that varied along the dimensions: size, material/color, and whether it was open or closed. Size competitors were just resized versions of the target. Open/Closed competitors were objects that were either identical to the target except for the relevant dimension, or matched as closely as possible.

Target	Material/Color Competitor
brown binder	blue binder
white shades	wooden shades
red book	brown book
wooden box	cardboard box
silver briefcase	black briefcase
red/yellow pill	gold/silver pill
silver cell phone	pink cell phone
brown casket	white casket
pink compact	metal compact
metal cabinet	wooden cabinet
disposable diaper	cloth diaper
paper fan	feather fan
metal filing cabinet	wooden filing cabinet
metal chair	wooden chair
real gun	plastic gun
metal hand cuffs	pink hand cuffs
glass bottle	metal bottle
purple knife	brown handled knife
metal stepladder	wooden stepladder
black piano	brown piano
metal pot	pot with glass lid
brown purse	white purse
silver refrigerator	white refrigerator
bronze scissors	plastic scissors
green teapot	white teapot
silver thermos	blue thermos
porcelain toilet	metal toilet
red metal toolbox	wooden toolbox
metal trashcan	plastic trashcan
white/green umbrella	red umbrella
black wallet	brown wallet
plastic zipper	metal zipper

References

Arts, A., Maes, A., Noordman, L., & Jansen, C. (2011). Overspecification facilitates object identification. *Journal of Pragmatics*, *43*, 361–374.

Baayen, R. H. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge University Press.

Bard, E. G., & Aylett, M. P. (2005). Referential form, word duration, and modeling the listener in spoken dialogue. In J. C. Trueswell & M. K. Tanenhaus (Eds.), *Approaches to studying world-situated language use*. MIT Press.

Barr, D. J. (2003). Paralinguistic correlates of conceptual structure. *Psychonomic Bulletin and Review*, *10*(462-467).

Barr, D. J. (2004). Establishing conventional communication systems: Is common knowledge necessary? *Cognitive Science*, *28*(6), 937–962.

Barr, D. J., & Kronmüller, E. (2007). Conversation as site of category learning and category use. In B. H. Ross & A. Markman (Eds.), *The Psychology of Learning and Motivation* (Vol. 47). Academic Press.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure in mixed-effects models: Keep it maximal. *Journal of Memory and Language*.

Barsalou, L. W. (1983). Ad hoc categories. *Memory & Cognition*, *11*(3), 211–227.

Barsalou, L. W. (1991). Deriving categories to achieve goals. In G. H. Bower (Ed.),

The Psychology of Learning and Motivation: Advances in Research and Theory (Vol. 27, pp. 1–64). San Diego, CA: Academic Press.

Bates, D., & Maechler, M. (2010). lme4: Linear mixed-effects models using S4 classes [Computer software manual].

Beun, R., & Cremers, A. H. M. (1998). Object reference in a shared domain of conversation. *Pragmatics & Cognition*, *6*(1/2), 121–152.

Bock, K., & Griffin, Z. M. (2000). The persistence of structual priming: Transient activations or implicit learning? *Journal of Experimental Psychology: General*, *129*(2), 177–192.

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

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

Brown-Schmidt, S., & Konopka, A. E. (2008). Little houses and casas pequenas: Message formulation and syntactic form in unscripted speech with speakers of English and Spanish. *Cognition*, *109*(2), 275–280.

Brown-Schmidt, S., & Tanenhaus, M. K. (2006). Watching the eyes when talking about size: An investigation of message formulation and utterance planning. *Journal of Memory and Language*, *54*, 592–609.

Carroll, J. M. (1980). Naming and describing in social communication. *Language and Speech*, *23*, 309–322.

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review*, *113*(2), 234–272.

Chang, F., Dell, G. S., Bock, K., & Griffin, Z. M. (2000). Structural priming as

implicit learning: A comparison of models of sentence production. *Journal of Psycholinguistic Research*, *29*(2), 217–229.

Clark, H. H. (1992). *Arenas of language use.* Chicago: University of Chicago Press.

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

Clark, H. H., & Krych, M. A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language*, *50*, 62–81.

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

Clark, H. H., & Murphy, G. L. (1982). Audience design in meaning and reference. In J. F. LeNy & W. Kintsch (Eds.), *Language and comprehension* (pp. 287–289). Amsterdam: North Holland.

Clark, H. H., Schreuder, R., & Buttrick, S. (1983). Common ground and the understanding of demonstrative reference. *Journal of Verbal Learning and Verbal Behavior*, *22*, 245–258.

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

Cleeremans, A., Destrebecqz, A., & Boyer, M. (1998). Implicit learning: news from the front. *TRENDS in Cognitive Sciences*, *2*(10), 406–416.

Crockford, C., Wittig, R. M., Mundry, R., & Zuberbühler, K. (2012). Wild Chimpanzees Inform Ignorant Group Members of Danger. *Current Biology*, *22*(1-5).

Deutsch, W., & Pechmann, T. (1982). Social interaction and the development of definite descriptions. *Cognition*, *11*, 159–184.

Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004). Perspective taking as egocentric anchoring and adjustment. *Journal of Personality and Social Psychology*,

$87(3)$, 327–339.

Ericsson, K. A. (2006). The influence of experience and deliberate practice on the development of superior expert performance. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge Handbook of Expertise and Expert Performance* (pp. 683–703).

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

Fincher-Kiefer, R. (2001). Perceptual components of situation models. *Memory & Cognition*, *29*(2), 336–343.

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

Gann, T. M. (2007). *Are Speakers Egomaniacs or Automatons? Evidence for Automatic Processing in Language Production.* (unpublished manuscript)

Gann, T. M., & Barr, D. J. (2012). Speaking From Experience: Audience Design as Expert Performance. *Language and Cognitive Processes.*

Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, 181–218.

Garrod, S., & Doherty, G. (1994). Conversation, co-ordination and convention: an empirical investigation of how groups establish linguistic conventions. *Cognition*, *53*, 181–215.

Garrod, S., & Pickering, M. J. (2004). Why is conversation so easy? *TRENDS in Cognitive Sciences*, *8*(1), 8–11.

Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language.* Oxford University Press.

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

Griffin, Z. M. (2004). Why look? Reasons for eye movements related to language production. In J. M. Henderson & F. Ferreira (Eds.), *The integration of language, vision, and action: Eye movements and the visual world* (pp. 213–247). Psychology Press.

Guhe, M. (2007). *Incremental conceptualization for language production.* Lawrence Erlbaum Associates.

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

Horton, W. S., & Gerrig, R. J. (2002). Speakers' experiences and audience design: knowing when and know how to adjust utterances to addressees. *Journal of Memory and Language*, *47*, 589–606.

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

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

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

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science*, *11*(1), 32–38.

Kirsh, D., & Maglio, P. (1994). On distinguishing epistemic from pragmatic action. *Cognitive Science*, *18*, 513–549.

Krauss, R. M., & Weinheimer, S. (1964). Changes in reference phrases as a function of frequency of usage in social interaction. *Psychonomic Science*, *77*, 622–626.

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

Levelt, W. (1989). *Speaking: From intention to articulation.* MIT Press.

Levelt, W., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1–75.

Lockridge, C. B., & Brennan, S. E. (2002). Addressees' needs influence speakers' early syntactic choices. *Psychonomic Bulletin and Review*, *9*(3), 550–557.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, *95*, 492–527.

Logan, G. D. (2002). An instance theory of attention and memory. *Psychological Review*, *109*(2), 376–400.

Mangold, R., & Pobel, R. (1988). Informativeness and instrumentality in referential communication. *Journal of Language and Social Psychology*, *7*(3-4), 181–191.

Markman, A. B., & Makin, V. S. (1998). Referential communication and category acquisition. *Journal of Experimental Psychology: General*, *127*(4), 331–354.

Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review*, *104*(2), 266-300.

Olson, D. R. (1970). Language and thought: Aspects of a cognitive theory of semantics. *Psychological Review*, *77*(4), 257–273.

Pechmann, T. (1989). Incremental speech production and referential overspecification. *Linguistics*, *27*, 89–110.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue.

Behavioral and Brain Sciences, *27*, 1–22.

Pickering, M. J., & Garrod, S. (2005). *Establishing and using routines during dialogue: Implications for psychology and linguistics.* Erlbaum.

Pickering, M. J., & Garrod, S. (2006). Alignment as the basis for successful communication. *Research on Language and Computation*, *4*, 203–228.

Proctor, R. W., & Vu, K. L. (2006). Laboratory Studies of Training, Skill Acquisition, and Retention of Performance. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge Handbook of Expertise and Expert Performance* (pp. 265–286). Cambridge University Press.

Rawson, K. A. (2010). Defining and investigating automaticity in reading comprehension. In J. P. Mestre & B. H. Ross (Eds.), *Psychology of learning and motivation* (Vol. 52, pp. 185–230). Academic Press.

Rehder, B., & Hoffman, A. B. (2005). Eyetracking and selective attention in category learning. *Cognitive Psychology*, *51*, 1–41.

Roßnagel, C. S. (2000). Cognitive load and perspective-taking: applying the automatic-controlled distinction to verbal communication. *European Journal of Social Psychology*, *30*, 429–445.

Roßnagel, C. S. (2004). Lost in thought: Cognitive load and the processing of addressee's feedback in verbal communication. *Experimental Psychology*, *51*(2), 1–10.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A Simplest Systematics for the Organization of Turn-taking For Conversation. *Language*, *50*(4), 696–735.

Saylor, M. M. (2000). Time-stability and adjective use by child and adult English speakers. *First Language*, *20*, 91–120.

Schegloff, E. A., Jefferson, G., & Sacks, H. (1977). The preference for Self-Correction in the Organization of Repair in Conversation. *Language*, *53*(2), 361–382.

Schober, M., & Clark, H. H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology*, *21*(2), 211–232.

Simon, H. A., & Chase, W. G. (1973). Skill in chess. *American Scientist*, *61*, 394–403.

Smith, C. S. (1961). A class of complex modifiers in English. *Language*, *37*(3), 342–365.

Tomasello, M. (1999). *The cultural origins of human cognition*. Harvard University Press.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Harvard University Press.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*, 352–373.

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

Wardlow Lane, L., Groisman, M., & Ferreira, V. S. (2006). Don't talk about pink elephants! Speakers' control over leaking private information during language production. *Psychological Science*, *17*(4), 273–277.

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

Wu, S., & Keysar, B. (2007a). The effect of culture on perspective taking. *Psychological Science*, *18*(7), 600–606.

Wu, S., & Keysar, B. (2007b). The effect of information overlap on communication effectiveness. *Cognitive Science*, *31*, 169–181.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, *123*(2), 162–185.