Stony Brook University

Assessing Common Ground in Conversation:

The Effect of Linguistic and Physical Co-Presence on Early Planning

A Dissertation Presented

by

Alexia Galati

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Experimental Psychology

Stony Brook University

December 2009

Stony Brook University

The Graduate School

Alexia Galati

We, the dissertation committee for the above candidate for the

Doctor of Philosophy degree,

hereby recommend acceptance of this dissertation.

Dr. Susan Brennan, Dissertation Advisor,

Professor, Psychology Department

Dr. Arthur Samuel, Chairperson of Defense,

Professor, Psychology Department

Dr. Hoi-Chung Leung,

Associate Professor, Psychology Department

Dr. Amanda Stent,

Principal Member Technical Staff, AT&T Labs-Research

This dissertation is accepted by the Graduate School.

Lawrence Martin

Dean of the Graduate School

Abstract of the Dissertation

Assessing Common Ground in Conversation:

The Effect of Linguistic and Physical Co-Presence on Early Planning

by

Alexia Galati

Doctor of Philosophy

in

Experimental Psychology

Stony Brook University

2009

Speakers routinely adjust their behavior upon assessing the information they share in common with their conversational partners, but there remains controversy over when and how these adjustments happen. In this dissertation I address two debates regarding partner-specific adjustments: (a) whether they recruit the language processing system in a way that is so automatic as to be inflexible, affecting more inferential processes (e.g., utterance planning) but not fast-acting ones (e.g. articulation), and (b) what aspects of an experience with a partner become indexed in episodic traces for shared information. Specifically, I investigate whether co-presence conditions (i.e., whether information is shared linguistically, physically, or both) become indexed in episodic traces, consequently affecting both utterance planning and articulation.

Experiments 1 and 2 involved referential communication tasks in which Directors instructed two Matchers, separately, on how to arrange cards. In Experiment 1 materials were items that were difficult to describe, whereas in Experiment 2 materials had

common labels. In the first two rounds (Phase 1) cards were distributed as follows: with each Matcher, some cards were shared linguistically and physically, others only linguistically, others only physically, and others were completely absent. In the subsequent two rounds (Phase 2) Directors matched all cards from Phase 1 with each Matcher. I examined whether the Directors' descriptions in Phase 2 reflected sensitivity to the co-presence conditions in Phase 1.

Indeed, Directors' initial descriptions in Experiment 1 showed sensitivity to how information had been shared. Directors used fewer definite expressions for items that had been mentioned in Phase 1 compared to items that had not been mentioned. At the same time, adjustments in the amount of detail and provisionality of their initial descriptions showed sensitivity to the specific conditions of co-presence, suggesting that episodic traces did not merely encode a binary (mentioned vs. unmentioned) distinction: Directors included more words, idea units, reconceptualizations and hedges for items they had shared only physically with their Matchers compared to items they had shared only linguistically, and in turn included more words, idea units, reconceptualizations and hedges for items they had shared only linguistically with their Matchers compared to items they had shared both linguistically and physically. When taken together, these adjustments reflect appropriate strategies in initial audience design, driven by speakers' memory for how information had been previously shared: Referents that had been shared both linguistically and physically involved attenuated initial descriptions (fewer words, idea units), fewer markers of provisionality and more markers of definiteness. Referents that had been shared only linguistically were described with just as many markers of definiteness, signaling to the conversational partner that these referents had been

previously mentioned. But at the same time they were described with more detail, with just as many words and idea units as completely new referents, reflecting the degree of grounding that the conversational partners had previously achieved.

Directors' explicit reports in a source monitoring questionnaire on how they had shared items with their Matchers in Phase 1 of Experiments 1 and 2 provide corroborating evidence that to some extent people can actually recall the conditions of co-presence.

The intelligibility of lexically identical expressions culled from Experiment 2 was assessed by a new group of listeners in Experiment 3. Listeners' judgments revealed that Directors also distinguished the intelligibility of their expressions according to how information had been shared: expressions for items that had been shared previously only physically were rated as clearer than those shared only linguistically or both linguistically and physically. In other words, although items in Experiment 1 lacked conventional labels and were negotiated more than those in Experiment 2, Directors kept track of co-presence and adjusted their utterance planning and articulation accordingly in both experiments.

Together, these findings suggest that episodic traces do index the conditions of co-presence of shared information: speakers' adjustments in utterance planning reflect grounding techniques appropriate to how information had been previously shared. Moreover, the effects of co-presence extend to the relatively automatic process of articulation, suggesting that partner-specific adjustments are deployed flexibly. When the informational needs of the conversational partner are represented easily and are cued

rapidly, as a small set or relevant constraints, speakers adjust their early planning at multiple grains of linguistic processing.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Acknowledgements

First, I'd like to thank Susan Brennan for the mentoring and support she has provided during my time at Stony Brook. Susan, you've always demonstrated tremendous enthusiasm and remarkable resourcefulness, and have taught me a great deal about the research process and writing. I'm indebted to you for all of this. By thanking you, I also want to extend my appreciation to NSF Grants IIS-0527585 and ITR-0325188 for supporting the research I've conducted at Stony Brook.

I'd like to thank Arty Samuel for his openness and his willingness not only to collaborate with me on a project but also to provide feedback on each of my other research endeavors. Arty, I really appreciate your methodical way of unpacking ideas, your availability, and the access you've given me to your lab's resources; these have been invaluable to my graduate training.

I'd like to thank the other members of my committee, Amanda Stent and Hoi-Chung Leung, for useful comments that have helped fine-tune the design and framing of this dissertation. I'd also like to thank Richard Gerrig for helpful discussions during the early stages of developing this project.

My exceptional undergraduate Research Assistants, Jenny Lee, Billy Galbavy, and Gennadiy Ryklin, also deserve major kudos for having helped me with many aspects of this project. All RAs have contributed to coding description onset times in Experiment 1, transcribing the category exemplar trials of Experiment 2, coding disfluencies for Experiments 1 and 2, and collecting data for Experiment 3. I would also like to thank Donna Kat for her technological assistance.

There are, of course, those outside academia who've been instrumental to keeping me steadfast on this journey. I would like to thank my dad, Michalis Galatis, for the encouragement he channeled through our comforting Sunday conversations about the simple things in life. Dad, you have helped ground my perspective onto what truly matters and have consistently injected a needed dose of equanimity in my life.

I would like to thank my mom, Anna Galati, for the support and inspiration she has always provided. Mom, you have set an exemplary precedent by leaving Cyprus at a young age to pursue your education abroad, despite your family's objections, and proceeding to become so accomplished in your field. Thank you for being the paragon of a working parent, and for instilling ambition and self-sufficiency in your only child so as to release her fearlessly into the world. Without your unconditional support—emotional, financial, and otherwise— none of this would be possible.

Finally, I would like to thank Vin Dang for just about everything. Vin, my graduate school years will be inextricably bound to memories of us in New York, so thank you for making the move. Thank you for enhancing every aspect of my life with your intelligence, joie de vivre and humor. And for your tech support, cultural attaché and design services (those impeccably selected fonts!). The lives of knowledge seekers are in flux for extended periods of time and I can only hope you'll be joining me on this adventure.

CHAPTER 1: Introduction

1.1. The memory-based view of audience design

When people try to achieve mutual understanding or a joint goal they routinely consider what information they have in *common ground* with their conversational partners –namely, what their mutual knowledge, goals, and beliefs are (Clark & Marshall, 1981; Stalknaker, 1974). Upon assessing what information they share in common with one another, people may adjust their behavior accordingly—a phenomenon characterized as *audience design*. Engaging in audience design means that speakers design their utterances by taking into account the needs of specific addressees, while addressees, in turn, interpret those utterances by taking into account that they are specifically intended for them (Clark & Murphy, 1982; Clark & Carlson, 1982).

There is little controversy over whether partner-specific adjustments following from audience design occur in conversation. Researchers have focused instead on the circumstances under which people are more or less likely to perform such partner-specific adjustments, investigating the effect of contextual, motivational, and individual factors. For example, the referring expressions people produce depend on how much experience they have had with their partners at establishing a joint perspective, through a process called *lexical entrainment* (Brennan & Clark, 1996). Speakers' ability to consider their partner's ease of understanding may also depend on their initiative to speak: when switching the roles of giving and following directions on a frequent (trial-by-trial) basis, speakers produced disambiguating cues only when the syntactic context was ambiguous (Haywood, Pickering, & Brannigan, 2005). In another study in which participants switched roles only once, speakers produced disambiguating cues less discriminately,

regardless of the ambiguity of context, since the status of information was not immediately available as being ambiguous to speakers: they took longer to compute the disambiguating information than it took them to begin to speak (Kraljic & Brennan, 2005). People's role in conversation also affects their understanding: overhearers find it more difficult than addressees to interpret referring expressions, presumably because overhearers are unable to coordinate reference with the speaker (Schober & Clark, 1989). Moreover, whether people have matching or discrepant goals when collaborating affects their contributions to the conversation and their memory for shared information (Wilkes-Gibbs, 1986; Russell & Schober, 1999). Even individual differences in working memory capacity can affect how quickly and accurately people coordinate perspectives with their partners, and what strategies they use (Lockridge, 2007).

In addition to these contextual, motivational and individual factors, audience design is influenced by the constraints of cognitive architecture. I endorse the view that the representations accessed during audience design are general-purpose episodic traces and the processes that act upon these representations work in a similar manner as in other memory-dependent processes, such as source memory and context-dependent memory (Horton & Gerrig, 2005a, 2005b). This view supplants early frameworks of audience design that have proposed specialized memory representations for partner-specific information (e.g., Clark & Marshall, 1981; Clark & Haviland, 1977). Such specialized memory representations are problematic because they involve great computational demands in updating moment-by-moment a model for the conversational partner's informational needs and raise difficulties in defining the boundedness of "shared events" (for a discussion see Horton & Gerrig, 2005a). On this view that audience design is

deployed through ordinary cognitive processes, for audience design to occur conversational partners must act as strong enough cues to make memory representations for shared information accessible and these representations must become accessible with the appropriate time course (see also Metzing & Brennan, 2003). Nevertheless, there remain some controversies and unanswered questions regarding the cognitive architecture and processes involved in audience design, which I address in the next section.

1.2. Controversies regarding audience design

While there is substantial empirical evidence that audience design does take place in conversation, there are two main controversies regarding the cognitive constraints underlying audience design. These involve: (a) the time course of audience design, and (b) the modularity of the language processing system; specifically, whether fast-acting processes (like articulation) are encapsulated from audience design. An additional unanswered question is what the limits are on what factors can be represented in the episodic traces for shared information that underlie audience design; specifically, what and how many relevant dimensions about the communicative situation these episodic traces can index.

With respect to the time course of audience design, one view is that early language processing gives priority to partner-specific information (a position attributed to Clark by Keysar, 1997), while a competing view is that early processing gives priority to egocentric information. This latter view of audience design is described by Keysar and colleagues as two-stage model (e.g., Horton & Keysar, 1996; Keysar, Barr, & Horton, 1998; Keysar, Barr, Balin, & Paek, 1998), which considers initial processing to be egocentric and partner-specific adjustments to follow late, in the form of repairs.

Language processing, according to the two-stage model, occurs initially without regard to the addressees' informational needs, defaulting to what is most immediately accessible (in line with demonstrations of availability-based effects during sentence production, e.g., Bock, 1986; V. Ferreira & Dell, 2000) while the products of initial processing are monitored (though the precise monitoring mechanisms remain unclear, as pointed out by Horton & Gerrig, 2005a and Polichak & Gerrig, 1998) and adjustments are made only when necessary. A view related to the two-stage model is one that assumes strictly that during initial processing people do not routinely assess memory representations for shared information and instead use simpler cognitive mechanisms, such as priming (Pickering & Garrod, 2004). According to Pickering and Garrod's (2004) interactive alignment model, conversational partners achieve converging mental representations without tracking anything specific about the information they have shared. Instead, they adapt to their partners by using their own knowledge as a proxy, which is presumably effective because, having evolved to have the same cognitive architecture, what is easier for speakers should also be easier for addressees (see also Brown & Dell, 1987).

The memory-based view can be seen as reconciling these two views regarding the time course of audience design. Since conversational partners serve as contextual cues for general-purpose episodic traces for shared information, partner-specific memory associations can have an early and immediate impact on utterance planning and processing when supported by the right set of cues. Indeed, partner-specific adjustments can be found early in processing, demonstrating agility in the language processing system (e.g., Hanna, Tannenhaus, Trueswell, 2004; Hanna & Tanenhaus, 2004, Nadig & Sedivy, 2004, Metzing & Brennan, 2003). At the same time, language processing is opportunistic,

using whatever information is most accessible within a given time frame, as described in constraint-based models of language processing (Tanenhaus & Trueswell, 1995; MacDonald, 1994; McRae, Spivey-Knowlton, & Tanenhaus, 1998). In constraint-based models, different information sources or constraints each contribute probabilistic evidence for alternative interpretations (and presumably also speech plans) that compete with one another. Therefore, if the right set of cues does not become available within the appropriate time course, then partner-specific associations may not be adequately activated and assessed to constrain processing and implement partner-specific adjustments early on. An implication here is that understanding the constraints of the time course of audience design involves, in part, understanding what the "right set of cues" is for making available representations of relevant aspects of shared information (an endeavor relevant to the issue of limits of representation described below).

The second controversy regarding audience design deals with the architecture of the language processing system, specifically in terms of whether the system engages in audience design in a modular fashion or not. Some researchers have suggested that some fast-acting and relatively automatic processes, such as articulation (during which phonetic plans are translated to motor commands), are encapsulated from partner-specific knowledge and default to being egocentric, whereas other processes, such as the planning of referring expressions, can be guided by inferences about a partner's needs (Bard et al., 2000; Bard & Aylett, 2001; Brown & Dell, 1987; Keysar, Barr, Balin, & Paek, 1998; Kronmuller & Barr, 2007). However, in Galati and Brennan (2010), we have not found support for such a "dual process model", having demonstrated instead partner-specific adjustments in both articulation and utterance planning. In our study, speakers

distinguished retellings to new and old addressees by adapting not only the amount of events, words, and details they incorporated in their stories, but also the intelligibility of their expressions. Nonetheless, it may be the case that when shared information is not represented simply or cued rapidly within the appropriate time course, speakers' fast-acting processes may particularly susceptible to egocentric processing, more so than inferential processes, and may under such circumstances fail to demonstrate partner-specific adjustments.

Finally, an aspect of audience design that is not well understood concerns the nature of the memory representations for shared information and the conditions under which conversational partners serve as stronger cues for them. Understanding how shared information is represented in memory and when it's most effectively cued can shed light on both the time course of audience design and the modularity of the language processing system: whether partner-specific adjustments are deployed early or late, and whether they are deployed both at the level of utterance planning and articulation may very well depend on the kinds of memory representations for shared information with which we are dealing. From a memory-based perspective, failures in audience design, such as misattributing information as having been shared with a particular person, could be thought of as reflecting general memory-dependent processes such as source monitoring difficulty (Johnson, Hashtroudi, & Lindsay, 1993). Consistent with this perspective, executive functioning–namely, failures in inhibiting one's own perspective when working memory is taxed— can account for some insensitivities to shared information (Brown-Schmidt, 2009b). Having the right set of cues within the right time frame, then, should allow speakers to overcome such difficulties and engage in partner-specific adjustments

early on and even in terms of articulation. Yet little research has elucidated what the right set of cues is for readily assessing relevant aspects of shared information represented in episodic traces. Investigating the circumstances under which conversational partners succeed in serving as effective cues for relevant aspects of shared information lends insight into what aspects of shared experiences with conversational partners are indexed in episodic memory traces.

So far, few studies have addressed *directly* what aspects of the unfolding shared experience with a conversational partner get encoded in episodic traces for shared information. There is evidence that episodic traces can include associations between conversational partners and category groupings for the shared information, as well as associations between conversational partners and expressions established through re-referring (Horton & Gerrig, 2005a, 2005b, 2002; Metzing & Brennan, 2003). When conversational partners are associated with category information (e.g., *I have talked to Ann about fish and birds and to Beth about turtles and dogs*), speakers are more likely to distinguish their referring expressions for shared and non-shared information with each partner than when these categorical distinctions are not available (Horton & Gerrig, 2005b, 2002). And when conversational partners are associated with entrained-upon expressions established through re-referring, people experience more interference when an old partner uses a new expression but not when a new partner does (Metzing & Brennan, 2003).

In addition to indexing what categories have been shared and what expressions have been entrained upon, conversational partners may index other dimensions of the shared experience in their episodic traces as well. In this dissertation, I investigate

specifically whether the conditions of co-presence under which information is shared with a partner—whether information is shared multimodally (linguistically and physically), unimodally (only linguistically, or only physically), or not at all—are indexed in episodic traces and can be easily cued to result in partner-specific adjustments at both the level of utterance planning and articulation. In other words, this dissertation addresses mainly the last two issues regarding audience design described above by investigating (a) when conversational partners serve as the most effective cues, lending insight into the underlying memory representations for shared information, and (b) whether partner-specific adjustments are deployed in a non-modular fashion throughout the language processing system or are instead limited to inferential processes.

1.3. Indexing how information has been shared

As discussed in the previous section, speakers can index what categories of information they've discussed with a partner as well as what expressions they've entrained upon (Horton & Gerrig, 2005a, 2005b, 2002, Metzing & Brennan, 2003). In fact, people can use a particular partner to cue both category and lexical associations, even in the absence of an explicit intent to communicate with them (Horton, 2007). But whether episodic memory traces encode *how* information is shared is still an open empirical question.

Increasingly, interacting with others over distance is enabled by multimedia technologies whose affordances—the possibilities for use, intervention and action they offer (Gibson, 1996)— vary in terms of several dimensions, including visibility, audibility, and simultaneity. In the collaborative framework of language use, proposed by Clark and colleagues, partners in a joint activity monitor and coordinate their behavior by

grounding, or exchanging ongoing evidence about what they do or do not understand (Clark & Brennan, 1991; Clark, 1996; Cahn & Brennan, 1999; Brennan, 1990, 2005). Moreover, when partners are engaged in grounding, they will adapt the techniques they use in order to minimize their collective effort, even if one partner, at some point, must put in more effort than the other; this is described as the principle of *least collaborative effort* (Clark & Wilkes-Gibbs, 1986). Therefore, assessing whether some information is in common ground may, under some circumstances, also require consideration of *how* that information came to be part of common ground: tracking how information has been shared may be important, since when the communicative situation misses certain affordances, people may have to put more time and effort into grounding their conversations.

For example, there is empirical support that how information is shared while conversational partners are collaborating affects the grounding techniques they use. Much of this work has examined whether grounding depends on whether the speaker has visual evidence of the addressee's understanding or not. In a study by Brennan (1990, 2005) conversational partners tried to align their icons on identical maps displayed on networked computer screens. When the person giving directions lacked visual evidence about their partner's icon movements, the pair went through a lengthier process of verbally checking that they mutually understood each other and the partner following directions was the one to propose when to move to the next trial once thinking they had understood well enough for current purposes. Similarly, being able to monitor the partner's workspace and actions improves the efficiency of coordination: in a study in which pairs built LEGO models together with one person giving instructions and the

other assembling, pairs for which the participant giving instructions could see the builder's workspace made fewer errors and were substantially faster, in part because they could precisely time their utterances to the actions the builders were performing (Clark & Krych, 2004). In a related study, Gergle, Kraut, and Fussell (2004) manipulated different features of shared visual space to assess whether it affected how conversational partners coordinated. They found that having a shared visual space, particularly in tasks that were dynamic (with objects that were changing and hard to describe), helped conversational partners understand better the current state of their task and enabled them to achieve better task performance and coordinate their conversations faster. These studies demonstrate that when people can rely on visual evidence, they achieve mutual understanding more efficiently, in part, because speakers are less likely to need to seek verbal evidence and addressees are less likely to provide it. For instance, when partners shared visual information, they were able to reduce collaborative effort both at the stage where speakers planned the utterance and at the stage where speakers assessed that their addressees had understood the utterance (Gergle, Kraut & Fussell, 2004).

In light of these convergent findings demonstrating that the affordances of the current communicative situation affect conversational partners' grounding techniques, I asked whether the affordances of a *prior* communicative situation are indexed in episodic traces for shared information and whether they affect the partners' current grounding techniques and the speakers' early choices in utterance planning. In this dissertation, I manipulated the affordances or co-presence conditions under which information has been shared. According to Clark and Marshall (1981), in ordinary conversation, awareness of the addressee's knowledge can come from three different sources of mutual knowledge or

common ground: *linguistic co-presence* (information that can be derived from past and present conversations between conversational partners), *physical co-presence* (the shared physical or perceptual environment of the conversational partners), and *community membership* (their shared sociocultural background). Here, I regarded information that has been shared through linguistic and physical co-presence to have been shared multimodally, and information that has been shared through only linguistic or only physical co-presence to have been shared unimodally. The central question of this dissertation is whether the way information had been previously shared gets indexed in memory representations of shared information, and whether that consequently affects partner-specific adjustments in early planning.

Let me exemplify these research questions with a real world scenario: Imagine three friends—Ann, Beth, and Charles— visiting a museum. On par with a typical museum experience, the three friends meander, intercepting at certain points in different wings of the museum to look at paintings together and sometimes talk about these paintings together. Some paintings Ann views and talks about with Beth; others with Charles. Some other paintings Ann and Beth just view together; others Ann and Charles just view together. And as they intercept each other along the hallways and stairways, Ann may also describe to Beth and Charles, at different points, some of the notable artwork she's seen in other wings. At a later point in time, imagine that Ann and Beth reminisce about their trip to the museum. Would it be easier for Ann to determine that she's shared a particular painting with Beth if it had been one that they had viewed and talked about together as opposed to one that Ann had described to Beth in passing or one that they had just viewed together? If so, Ann may be more likely to mark paintings they

had shared multimodally as shared, for instance, through a definite expression and using a succinct description (e.g. “I really liked the Blue Rider painting”), while she may introduce a painting they shared only physically more tentatively, using hedging expressions and adding more detail (e.g., “I really liked a kind of Miro-looking, Bauhaus painting by Kandinsky.”).

From the perspective of constraint-based models of language processing, what Ann had shared with Beth at the museum –their common ground—can be thought of as serving as a contextual constraint that has immediate and probabilistic effects on Beth’s interpretation and presumably also on Ann’s utterance planning during their later interaction, depending on the strength and salience of Ann’s perspective and its relevance to the Beth’s informational needs (Hanna, Tannenhaus, & Trueswell, 2004). This approach predicts that the effects of shared information will depend on its strength and saliency, and its relevance to the task at hand. If those paintings that Ann had discussed and viewed jointly with Beth are represented more saliently in memory compared to those that she had only discussed or only viewed with her, her adjustments in her subsequent descriptions should reflect this.

In line with this approach we have proposed, practically and not strictly, that when information about the partner’s informational needs is already available or can be cued rapidly and unambiguously, speakers appear to represent and use simple distinctions about the partner’s needs at little or no discernable computational cost (Galati & Brennan, 2010; Brennan & Hanna, 2008). Indeed, in a number of experimental studies in which the situation permits a binary, either/or distinction about the partner’s informational needs, speakers can make adjustments specific to their partners’ informational needs early and

not in the form of repairs, reflecting that the partner's informational needs serves as a binary "one-bit" constraint. For example, speakers' linguistic behavior can distinguish instances when the partner is a native speaker of the same language or not (Bortfeld & Brennan, 1997), when the partner is an expert or not (Isaacs & Clark, 1987), when the partner has a picture of what is being described or not (Lockridge & Brennan, 2002), when the partner has referred to a particular object before or not (Metzing & Brennan, 2003), when the partner has heard the story before or not (Galati & Brennan, 2010), or when the partner can see the speaker or not (Bavelas et al, 1995; Cohen & Harrisson, 1973).

But many communicative situations, including Ann and Beth's interaction following their museum visit, may have more than a single constraint at work: namely, not just whether some information was previously discussed or not, but also whether it was shared physically together or not. Constraint-based models of language processing can allow multiple constraints about common ground to be used; whether these constraints in fact influence processing should depend on the availability of information, given the limitations of working memory, and their relevance to the task.

The prediction that sharing information multimodally will lead to stronger episodic traces than sharing information unimodally is consistent with findings that people learn and remember information they have experienced multimodally better than when they have experienced it in one modality (see Shams & Seitz, 2008 for a review). For example, seeing accompanying pictures improves the memory for words (Anderson & Bower, 1973; Paivio, 1986) and the learning of text (Carney & Levin, 2002). Moreover, there is evidence that stimuli with multisensory pasts are more accurately

discriminated as having already been seen: people can differentiate new from already viewed images according to whether their prior presentation was only visual or in an auditory-visual pair (Murray, Michel, Grave de Peralta Menendez, Ortigue, Brunet, Gonzalez Andino, & Schnider, 2004). Through the same process, experiencing shared information multimodally may lead to stronger memory representations as opposed to experiencing it unimodally. A relevant finding here is that repetition may reinforce the memory for shared information, as there is evidence that reusing an expression to refer to an object in a particular way increases the probability of continuing to refer to that object in the same way. Brennan and Clark (1996) found that, in a study in which pairs of participants worked on a card matching task, whether speakers used a basic level term (e.g., *the dog*) or more specific term (e.g., *the golden retriever*) did not depend on the most recent term they had used, but rather on the context of other cards and, critically, on the number of times they had previously referred to a card using a given term. In other words, repetition reinforced how effectively conversational partners served as cues for memory representations of shared information. By the same token, the multimodality of how information has been shared may also reinforce how effectively conversational partners serve as cues.

Of course, given the limitations of working memory, an infinite (or even a large) number of distinctions or constraints about common ground cannot be at work during language processing. Instead, constraints about common ground, like any other type of constraint (including discourse context and within-sentence structural and lexical biases), should be weighted depending on their availability and relevance to the task and be integrated with each other in parallel. For example, in the museum scenario, if the fact

that a painting had been previously shared vs. not (or else, discussed vs. not) is what is most salient and relevant, then Ann can compute this into a binary, "one bit" distinction: a simple constraint that could be rapidly cued to lead to partner-specific adjustments from the earliest moments of processing. But, if the specific conditions of co-presence are salient and available in working memory and are relevant to the task at hand (and the studies reviewed earlier—Brennan, 1990, 2005; Clark & Krych, 2004; Gergle, Kraut & Fussell, 2004—highlight why they may be relevant in terms of grounding), she may instead recruit two constraints: one about whether information had been previously discussed vs. not and another about whether information had been previously shared visually vs. not[1]. Whether speakers can recruit two constraints (or a non-binary constraint) about shared information from the earliest moments of processing has not yet been empirically demonstrated, but such constraints may still constitute a simple enough set of distinctions to be easily computed and have an immediate impact in processing. Critically, if shared information is strengthened probabilistically through the conditions of co-presence and these distinctions are available and relevant, then any resulting partner-specific adjustments should vary according to the conditions of co-presence.

1.4. The current project

[1] An alternative to having two constraints is that there may be a single constraint with a tertiary or quarternary distinction of the types of co-presence: shared linguistically and physically, shared only linguistically, shared only physically, and not shared at all.

In this dissertation I investigate whether the conditions of co-presence get indexed saliently enough in memory representations for shared information to lead to early partner-specific adjustments.

Experiments 1 and 2 parallel the circumstances of the museum example: the conditions of co-presence for shared information are manipulated to investigate whether partner-specific adjustments in speakers' early planning vary as a function of whether information was previously shared both linguistically and physically, only linguistically, only physically, or is completely new to the partner. Experiment 1 examines whether partner-specific adjustments at the level of utterance planning demonstrate sensitivity to how information had been previously shared, by considering measures that capture the amount of propositional content encoded, reconceptualizations, provisionality, and the ease of assessing shared information. Experiment 2 examines whether partner-specific adjustments at the level of articulation demonstrate sensitivity for how information had been shared; the intelligibility of lexically identical expressions assessed in Experiment 3. These experiments together shed light both on the architecture of the language processing system and on the limits of the underlying memory representations guiding partner-specific adjustments.

CHAPTER 2: General Method of Experiments 1 and 2

Experiments 1 and 2 address whether speakers' partner-specific adjustments depend on how they have previously shared this information with their partners. Specifically, these experiments attempt to elucidate whether the memory representations for shared information index the way information has been previously shared and whether these indexed distinctions result in partner-specific adjustments across levels of linguistic processing.

These experiments place participants in a referential communication task (an experimental paradigm introduced by Krauss & Weinheimer, 1966) in which one participant has information that the other requires. The person given primary responsibility to direct their partner has generally been called the Director and the other person has been called the Matcher. I have modified the referential communication task such that some items are shared both linguistically and physically between a Director and a Matcher, while others are shared only linguistically or only physically.

When Directors assess episodic traces of shared information, if the modalities in which information was shared become available as a single binary cue or constraint, then the Matchers' identity serves as a cue for whether the information has shared or not shared (or had been mentioned or not mentioned), but not for the specifics of how it has been shared. On the other hand, if the modalities in which information was shared become available as a small set of cues or constraints indexing how the information was shared, then Matchers should cue the specific conditions of co-presence, with information shared multimodally being represented more strongly than information shared unimodally or not at all. In other words, I ask whether the constraints provided by the

Matchers' identity would cue the Directors solely about whether the information was shared (or mentioned), or about specifically *how* information was shared (linguistically, physically, or both).

2.1. Design

In both Experiments 1 and 2, Directors had a stack of cards in prearranged orders (*target orders*), which they matched with two Matchers (A and B) separately. Each Matcher had a set of cards spread out in front of them in their *staging area*, which was occluded from the Director's view by a barrier. The primary goal of the task was for the Matcher to place their cards on the *target area*, a board with numbered slots, according to the Director's target order. The target area was visible to both the Director and the Matcher.

This referential communication task included two phases: the first served to establish shared information under different conditions of co-presence, and the second to assess whether Directors modified their linguistic behavior depending on how they had shared items in Phase 1.

The task involved matching sets of eight target cards. In Phase 1, the Director matched a *subset* of the eight target cards with the two Matchers (A and B) separately, in two rounds. To give rise to the two co-presence conditions for which items are shared only linguistically and only physically, the Director and the Matcher of a given round of Phase 1 did not have identical sets of cards. Specifically, of the eight target cards, two existed in both the Director's and Matcher's sets, two existed in only the Director's, two in only the Matcher's, and the two remaining cards were not shared at all in Phase 1. Table 1 illustrates the basic design of Experiments 1 and 2.

An important aspect of the design is that the distribution of co-presence conditions for each of the target cards between Matcher A and B, in Rounds 1 and 2, was complementary such that: (1) items that were shared linguistically and physically by the Director and Matcher A were completely absent for Matcher B, (2) items were shared linguistically by the Director and Matcher A were shared physically by the Director and Matcher B, (3) items that were shared physically by the Director and Matcher A were shared linguistically by the Director and Matcher B, and (4) items that were completely absent for the Director and Matcher A were shared linguistically and physically by the Director and Matcher B. Critically, all items had been described only once, to either Matcher A or B, in Phase 1. Having all items be linguistically "given" for the Director by the end of Phase 1 was necessary in order to avoid confounding adjustments stemming from what is easiest for the Director to do (e.g., attenuating a description of a referent that she had described before) with adjustments tailored to the current Matcher's informational needs (e.g., attenuating a description of a referent that the current Matcher had heard or seen before) (Keysar, 1997).

To handle the matching of the cards that were unique to the Director or Matcher's sets in Phase 1, in conjunction with the main goal of the referential communication task (to match the cards in the target order), two additional rules were introduced: (1) When the Director had a card that the Matcher hadn't, the Matcher used a Joker (placeholder) card. (2) When the Matcher had a card that Director hadn't, at the end of the round the Matcher handed the card to Director to confirm, without talking, that she indeed did not have it and place it at the end of the card sequence on the target area.

These additional rules of Phase 1 enabled me to establish different conditions of co-presence for different items. By the end of a round, the 6-slot board of Phase 1 had been filled with two cards that had been shared linguistically and physically by the Director and Matcher, two cards that had been shared only linguistically by the Director and Matcher (with joker cards standing in for them), and two cards, in the final slots, 5 and 6, that have been shared only physically (and had never been discussed) by the Director and Matcher. The Matcher had to coordinate with the Director to determine whether the Director was describing a card that the Matcher had or not (particularly with the materials of Experiment 1). Pairs could interact freely (Matchers could ask Directors clarification questions) and were allowed to correct any errors (as when the Matcher put up a card that wasn't the one that the Director described), and could even backtrack to correct an earlier error (as when the Matcher ran out of joker cards and realized that he must have misplaced a joker card for a card he had in an earlier position). As a result Directors and Matchers would always match correctly the first four cards (two of which the Matcher had, two of which he hadn't); the remaining cards in the Matcher's staging area were then put by the Director on the target area without being discussed—these were shared only physically.

In Phase 2, the Director matched all eight target cards with each Matcher separately in two more rounds. For each Matcher in Phase 2, a quarter of the items had been previously both linguistically and physically co-present, another quarter of the items had been previously linguistically but not physically co-present, a third quarter of the items had been previously physically but not linguistically co-present, and a final quarter of the items was completely new (had been neither physically nor linguistically co-

present). As with Phase 1, pairs could interact freely and correct any errors. The matching in Phase 2 was done in one of two orders: either Matcher A returned first for Phase 2 followed by Matcher B (Matcher order ABAB), or Matcher B remained for Phase 2 and Matcher A returned after that (Matcher order ABBA). This aspect of the design allowed me to examine whether interference from interacting with another partner attenuates partner-specific adjustments, by looking for differences in the adjustments made to Matcher B in the two orders.

The two experiments were conducted following this two-phase design, examining audience design effects at different levels of linguistic representation. In Experiment 1, I investigated whether partner-specific adjustments at the level of encoding propositional content reflected sensitivity to how the information had been previously shared. Abstract geometric shapes called *tangrams* were used as materials. I assessed partner-specific adjustments in terms of the number of words, idea units, reconceptualizations, and hedging in the Director's initial descriptions in Phase 2, as well as in the number of disfluencies and the amount of time taken to produce the first content word of the description (description onset time).

In Experiment 2, I investigated whether partner-specific adjustments in articulation reflected sensitivity to how information had been shared. Line drawings of common objects from different categories were used as materials. Triplets of lexically identical expressions from Phase 1 and Phase 2 (Rounds 3 and 4) were excised and their intelligibility was assessed in a separate experiment (Experiment 3).

Directors and Matchers participated in both Experiments 1 and 2, with the order of the experiments counterbalanced. Moreover, each item for each experiment was

presented in one of four item lists, such that across triads of participants it occurred equally frequently in each of the four co-presence conditions and in each of the eight target area positions (slots) in Rounds 3 and 4.

2.2. Participants

Ninety-six students from Stony Brook University participated in triads. Thirty-two of the participants served in the role of Director, and 64 served as Matchers. Participants were assigned to their roles randomly, by drawing cards. All Directors and Matchers were native speakers of English. Fifty-eight of the participants were female, 38 were male. Participants received research credit toward a requirement for a psychology course or prorated payment at the rate of $9/hour.

2.3. Materials

The materials (tangrams and line drawings of common objects) used in Experiment 1 and 2 are described in detail in Chapters 3 and 5, respectively.

2.4. Apparatus

Two boards were constructed to serve as the target area onto which cards were placed in Phases 1 and 2. The target area for Phase 1 was a 12 by 36 inch board with six slots, numbered 1 through 6. Slots 5 and 6 of this board (onto which Directors placed the physically co-present cards towards the end of a round in Phase 1) were highlighted in yellow. The target area for Phase 2 was a 12 by 40 inch board with 8 slots, numbered 1 through 8. The Director and Matcher in a given round sat at the sides of a table facing one another but separated by a barrier that occluded the Matcher's staging area and the Director's stack of cards (but not each other's faces). On the Director's right and Matcher's left was the target area, which they could both see and the Matcher could

easily reach. The recording equipment was staged as follows: one digital camcorder was placed behind the Director, filming a view of the Director's cards and perspective of the Matcher and target area; another digital camcorder was placed in front of the target area providing a clear view of it and side views of the Director and Matcher; a high quality audio recorder receiving input from the Director's headset was also placed behind the Director. Figure 1 illustrates the arrangement of Directors and Matchers in Experiments 1 and 2.

2.5. Procedure

All participants were told that the study investigated collaboration. They drew cards to determine their roles as Director, Matcher A, and Matcher B. The "rules" of the matching task were then explained. Participants were told that they would be playing with four different sets of cards, some of which would be pictures of abstract geometric shapes and others would be pictures of common objects. They were told that for each set of cards, they would be playing for four rounds, that in the first two rounds the Director would be matching a subset of the cards with Matcher A and B separately and that special rules for these two rounds applied since the Director and Matcher would not have the same set of cards. The special rules of Phase 1 were specified: (1) when the Director had a card that the Matcher didn't, the Matcher would be using a placeholder, "joker" card, and (2) after matching all four of the Director's cards, the Matcher would give the Director his remaining cards, the Director would inspect them to confirm that indeed she did not have them and would silently place them on the board in slots 5 and 6. They were then told, that in the last two rounds of playing with a given set of cards (Rounds 3 and 4), they would be matching some cards from the previous rounds, but that now both the

Director and Matchers would have the same set of cards and the rules of the first two rounds would no longer apply. They were told that, although Matchers would have the same eight cards as the Director, they would also have an extra, ninth card; this was to make it necessary for the Director to describe the final, eighth card. Directors were also instructed not to turn over the next card in their stack until the Matcher had placed the current card on the target area. Directors and Matchers were told that they could correct any errors they made—for instance, if the Matcher placed the wrong card on the board, the Director could correct them. Or in an alternative scenario, if the Matcher was out of joker cards, and the pair was stuck on a card that the Matcher did not seem to have, they could backtrack and revisit earlier cards to determine whether a joker card had been mistakenly placed on the board. All participants were asked to ignore the experimenter, who remained in the room in order to turn the recording equipment on and off between rounds, give the Director and Matchers their cards for each round, and change the target area boards between Phases 1 and 2.

Upon being given instructions, the Director completed two Phase 1 practice rounds with tangram cards, with each of their Matchers. Then the Director and Matchers participated in Experiment 1 and Experiment 2, and their exchanges were video and audio-recorded. During the experiment, when the Director matched cards with one Matcher, the other Matcher waited outside the room for their turn.

After participating in Experiments 1 and 2, the Director and Matchers were administered source monitoring questionnaires that asked how each item from Experiments 1 and 2 had been shared with their partners in Phase 1. Participants were then debriefed and paid, if participating for payment.

2.6. Analyses

Unless indicated otherwise, analyses were conducted as 2 x 2 x 4 ANOVAs with Matcher identity (Matcher A vs. Matcher B), Matcher order (ABAB vs. ABBA), and co-presence status (shared linguistically and physically, shared linguistically only, shared physically only, completely new) all as within-subjects factors. For each result, I report two analyses: *F1* is the analysis by subjects (for which means are computed for triads of participants), *F2* is the analysis by-items (for which means are computed for tangram cards, category exemplar cards, or excised expressions, as relevant). Effects of the order of Experiments 1 and 2, the set of tangram cards or exemplar cards, and the item list are not reported, as preliminary analyses revealed that the effects of these factors were not significant.

CHAPTER 3: Experiment 1 (Tangrams) Materials and Coding

With the objective of examining partner-specific adjustments at the level of encoding propositional content, in Experiment 1 I used materials that required some effort to lexicalize and some coordination between the Director and the Matcher, with the Matcher having to ratify the Director's proposal for a perspective and the Director ultimately confirming or grounding a match. In the Director's initial description in Phase 2, I examined: (a) the occurrence of definite expressions and meta-references, (b) the number of words, (c) the number of idea units, (d) the number of reconceptualizations, and (e) the frequency of hedging and metacomments. I reasoned that these measures could reflect partner-specific adjustments in terms of encoding content that would reflect sensitivity to how the information was shared in Phase 1.

I also examined two more measures that I reasoned would reflect how easily shared information is accessed in Phase 2: (a) the number of disfluencies preceding the first content words of a Phase 2 description and (b) the amount of time it took Directors to produce the first content words of a Phase 2 description.

Moreover, I examined some additional measures that do not reflect partner-specific adjustments, but rather reflect differences in coordination between partners in Phase 2 based on how information had been shared in Phase 1. These measures are: (a) the number of conversational turns for a description in Phase 2, (b) the type of the Matcher's contribution for a given description in Phase 2, and (c) the number of errors for a description in Phase 2.

In addition to examining exchanges in Phase 2, I coded for the total number of words and idea units, total amount of hedging, number of turns and the type of the

Matcher's contribution for descriptions in Phase 1. This was done to establish whether Directors and Matchers in Phase 1 had a different experience grounding cards that were linguistically and physically shared compared cards that were shared only linguistically.

Finally, the explicit memory of Directors for how they had shared items with each of their Matchers in Phase 1 was also assessed through the source monitoring questionnaire administered at the end of the session.

3.1. Materials

3.1.1. Stimuli. Two identical sets of 16 tangrams (pictures of abstract geometric shapes) were printed on 3 x 5 inch index cards. The tangrams were grouped into two sets of eight cards: one set resembling human figures, the other resembling animals (see Appendix A for a list of these two sets). Directors and Matchers used each of these sets when playing for two separate sets of four rounds. For each of the two sets, an additional tangram (a human-like figure and an animal-like figure) formed the Matchers' ninth card in Phase 2. The tangrams were selected to have comparable naming times according to results from a norming study by Swets, Jacovina, and Gerrig (in revision). Two additional images of jokers were printed on 3 x 5 inch index cards to be used by Matchers in Phase 1.

3.1.2. Source monitoring questionnaire. The questionnaire consisted of four pages, each page displaying one set of the cards matched in the experiment (two for the tangrams, two for the category exemplars). The order of the pages corresponded to the order in which participants had played with these sets of cards in their session. The pages for the tangram sets listed pictures for each of the 16 target tangrams. The Directors' questionnaire included two columns next to each tangram picture, one with a heading that

asked them how they experienced each card with Matcher A in Phase 1 (when playing with the 6-slot board), and a second with the corresponding heading for Matcher B. For each tangram, Directors had to chose one of the following options with respect to Matcher A: (a) both I and Matcher A had it, (b) only I had it (Matcher A used a joker), (c) only Matcher A had it (I placed it in slot 5 or 6), (d) neither I nor Matcher A had it, (e) I don't remember. The corresponding options with respect to Matcher B were listed for each item on the second column. Only the Directors' questionnaires were of interest here. Matchers were also given an analogous questionnaire, in which they indicated how they had shared each item with their Director, but their data were not analyzed[2].

3.2. Transcription

For each triad, all four rounds of both sets of tangrams were transcribed in detail, including contributions by both the Matcher and the Director. The transcripts included annotations of fillers such as "uh" or "um", pauses, lengthening of vowels, interruptions (both self-interruptions and interruptions by the Matcher) and restarts. Instrumental actions performed by the participants (such as the Matcher placing a card on the board) and non-verbal feedback, such as head nods and facial displays (such as frowning to indicate confusion), were also annotated in the transcripts.

3.3. Coding

The first five measures (definite expressions, number of words, number of idea units, number of reconceptualizations, and hedging), involved considering the Director's *initial descriptions* of tangrams in Phase 2. Initial descriptions were defined as the

[2] The first four triads were not administered source monitoring questionnaires. The source monitoring results in Chapters 4 and 6, for Experiments 1 and 2 respectively, involve analyses of data from 28 Directors.

Director's first turn, before any explicit feedback (i.e. an overt contribution) from the Matcher[3]. By focusing on this first contribution I aimed to tap onto partner-specific adjustments that were *not* performed as repairs, in response to the Matcher's feedback, but rather were part of initial planning. Of course, this way of defining the initial description did not preclude the fact that Directors during this time could have still been monitoring their Matcher's non-verbal behavior (e.g., continuing to survey their cards) and may have adjusted their linguistic behavior accordingly, as a late response. To address this caveat, I also coded two additional measures that may reflect initial audience design, independent of feedback of the Matcher: the number of disfluencies preceding the first content word of the description and the amount of time to produce that first content word. The remaining measures derived from transcripts (number of turns, the type of the Matcher's contributions, number of errors) took as a unit of analysis the entire exchange between the Director and Matcher for a given tangram, and were indicative of the coordination between pairs as opposed to initial audience design.

3.3.1. Definite expressions and meta-references. Speakers are more likely to mark explicitly that something is shared (e.g. through a meta-reference like the phrase

[3] There were two exceptions to this definition of the Directors' initial descriptions, pertaining to cases where the Director's first turn did not constitute their full initial description because the Matcher made a bid for the conversational floor. One exception involved cases where the Matcher attempted to speak during the Director's first turn, but the Director continued speaking without pausing or interrupting herself (typically in these cases Matchers aborted their bid for a turn almost immediately). The other exception was when Matcher provided feedback (e.g., saying "mm-hmm" or laughing) while the Director spoke without pausing or interrupting herself in response to this feedback. In both of these situations, I interpreted the Directors' continued speaking, despite feedback or attempted bids for a conversational turn from their Matchers, to indicate that they had not yet finished with their initial description. In these cases I took the Matcher's second contribution (as opposed to their first, which interrupted or overlapped with the Director's speech) to mark the end of the Director's initial description.

"*from the last time*") the more frequently they have referred to a particular object with their partner (Brennan & Clark, 1996). Here, I investigated whether the conditions of co-presence in Phase 1 affected how likely Directors were to mark explicitly a referent as shared.

I coded for the presence of definite expressions (e.g. "*the person falling*", "*the trumpet player*", "*the one with the stretched out arms, and the square triangular shoulders, and the long head*"), noun phrases preceded by definite pronouns (e.g., "*our clown*", "*our, like, really fat guy*", "*our morphed creature*"), and meta-references (e.g., "*it's already come up before*", "*I think I described this one to you before*", "*…that you guessed from the last time*", "*if you remember from last time*", "*that duck character we talked about earlier*", "*this looks like a person holding a machine gun, again*").

Definite expressions that arose from community membership, as opposed to linguistic or physical co-presence were excluded from coding. These included introductory mentions (typically in Phase 1) with definite expressions like "*the batman symbol*", "*the letter M/Y/I*", "*a guy giving the Nazi salute*", "*the presidential seal*", "*the geese that you see on campus*", or "*those monkey people with those pointy hats in the Wizard of Oz*". In these cases, Directors made inferences about historical, pop-cultural, graphemic, or campus-related knowledge they shared with their Matchers based on their shared socio-cultural background as opposed to the co-presence conditions of Phase 1.

3.3.2. Number of words. Speakers use more words when conveying information that is new to a conversational partner than information that is old (Galati & Brennan, 2010). Here, I wanted to replicate this finding and examine whether the co-presence conditions of sharing information are distinguished through word counts as well. Word

counts of the Director's initial description (their first turn, preceding any explicit feedback from the Matcher) were used as a first approximation of the informativeness of each tangram description in Phase 2. The Directors' total number of words for descriptions in Phase 1 was also determined. Contracted words (e.g., not, is, are, have), audible feedback responses (*mm-hmm*, *nuh-uh*) and expressive exclamations (e.g., *wow*, *oh*, *ugh*) were counted as separate words. Words interrupted mid-word were excluded from the word count.

3.3.3. Idea units. Although word counts could generally indicate whether Directors attenuated or elaborated their descriptions, they may not have always been a precise indicator of such adjustments because words don't map directly onto propositional content, as they also include hedging and metacomments (see Sections 3.3.5 and 3.3.6). Coding for the idea units that the Directors produced during their first turn provides a more accurate measure of partner-specific adjustments in encoding propositional content. This can extend previous findings that speakers include more idea units when items are new for a partner than when they are old, as long as they can keep track of what categories of items they've previously talked about (Horton & Gerrig, 2002). In addition to the initial idea units of descriptions in Phase 2, I also determined the total number of idea units of descriptions in Phase 1.

Similar to Horton and Gerrig (2002), I coded each content word—each noun, adjective, action verb, adverb, and prepositional phrase—as belonging to a separate idea unit. Verb phrases that suggested directionality (e.g., "*facing to the left*", "*pointing to the left*") were counted as single idea units just like prepositional phrases (e.g., "*to the left*"). Compound nouns (e.g., "*stop sign*", "*lava lamp*", "*bicycle kick*", *"negative space*",

"*muscle man*") were counted as a single idea unit. Appendix B shows an example from Triad 30 of how the initial and total number of idea units in a description from Phase 1 were identified.

3.3.4. Reconceptualizations. Speakers use more reconceptualizations with a new partner than with an old partner when they can keep track of what categories of information they've previously talked about (e.g. Horton & Gerrig, 2005b, 2002), which is consistent with the idea that reusing a perspective that has been established with a specific partner should help that partner readily identify the intended referent (Brennan & Clark, 1996; Metzing & Brennan, 2003). Here, I'm examining whether Directors' use of reconceptualizations in their initial description decreases when they had previously shared an item through more conditions of co-presence.

To determine the number of reconceptualizations I considered the idea units in the initial description to a Matcher in Phase 2 and the total set of idea units the Director produced in Phase 1. I then determined which of the idea units in the initial description in Phase 2 overlapped with those of Phase 1, and which were unique to Phase 2. The number of idea units that were unique to Phase 2 was considered to be the number of reconceptualizations. Identical idea units were classified as overlapping, but synonyms or expressions that were semantically equivalent (e.g., *large-big*, *somebody-person*, *sideways-to the side*, *facing to the left-looking to the left*) were also taken to capture overlapping idea units.

Recall that every item had been described once in Phase 1 and that across Matchers the co-presence status of a given item was complementary (i.e., items that were shared linguistically and physically with Matcher A were absent in the round with

Matcher B, and vice-versa, and items that were shared only linguistically with Matcher A were shared only physically with Matcher B, and vice-versa). Therefore, assessing reconceptualizations in Phase 2 for a tangram that had been shared only physically with a Matcher in Phase 1 involved comparing the initial idea units for that tangram in Phase 2 to the total idea units of the description to the other Matcher (who didn't have it) in Phase 1. Likewise, assessing reconceptualizations for a tangram that was new for a Matcher in Phase 2 involved comparing the initial idea units of that description to the total idea units of the description to the other Matcher (who did in fact have it) in Phase 1. Assessing reconceptualizations for tangrams that had been shared linguistically and physically or only linguistically involved comparing, for the same Matcher, the idea units of the initial description in Phase 2 to the total number of idea units of the Phase 1 description. Appendix C illustrates how the reconceptualizations of Triad 30 were identified for both Matchers in Phase 2.

3.3.5. Hedges. The less certain speakers are that their partners will accept how they conceptualize a referent, the more likely they are to mark their description as provisional through hedging (Brennan & Clark, 1996; Horton & Gerrig, 2002). The following kinds of expressions were counted as hedges: kind of, maybe, probably, possibly, almost, in a way, in some sense, a little bit, if anything, -looking thing (e.g., "*shield-looking thing*"), -type thing (e.g., "*foot-type thing*"), -ish (e.g., "*triangle-ish*", "*back neck-ish*"), -y (e.g., "*rhombus-y*"). Hedges were identified in initial descriptions in Phase 2, and in the complete descriptions in Phase 1.

3.3.6. Metacomments. Another class of expressions that I identified as marking provisionality in a way similar to hedges was metacomments. Metacomments were

expressions that were not part of the tangram description but were instead editing expressions regarding the Director's perspective, commitment toward what they were saying, or difficulty in describing the card (e.g., "*I don't know*", "*I guess*", "*I think*", "*I'd say*", "*I assume that is*", "*that's what it looks like to me*", "*if you use your imagination*", "*if you look at it like that*"), as well as emotional reactions in response to the card (e.g., "*oh my god!*", "*oh my!*", "*oh no!*", "*oh, what is this?*", "*wow!*", "*ugh!*", "*whew!*"). A final type of metacomment I coded for involved the Director's explicit questioning of the Matcher's knowledge of a specific expression or domain (e.g., "*do you know what a bicycle kick is*?", "*do you know biology? enzymes*?"). Metacomments were identified in initial descriptions in Phase 2, and in the complete descriptions of Phase 1.

Note that both meta-references and meta-comments are expressions that are not part of the tangram description, but rather pertain to the meta-narrative level: either the pair's prior experience or the Director's or Matcher's mental state. Through meta-references Directors invoke their prior experience with their Matcher (e.g., *it's already come up before*) and in this sense these expressions are related to definite expressions, while through metacomments Directors invoke their own commitment or difficulty in describing a card and in this sense these expressions similar to other markers of provisionality, like hedges.

3.3.7. Disfluencies. Since disfluencies can signal the information status of a referring expression (see Arnold, Tanenhaus, Altmann, & Fagnano, 2004; Brennan & Schober, 2001) or the speaker's metacognitive state, such as their commitment to their response (Brennan & Williams, 1995), they may also exhibit sensitivity to how information was shared with a Matcher. Therefore, the total number of disfluencies

preceding the first content word (most often noun or an adjective) that was part of the description of the current card was determined. Two types of disfluencies were considered here: fillers like "uh" and "um" and self-interruptions. Disfluencies preceding the first content word were identified only for descriptions in Phase 2. For example, the initial description "*and this one looks like* / like <um> a flamingo*" contains one filler and one self-interruption before the first content word.

3.3.8. Description onset time. Earlier description onset reflects greater ease at establishing whether an item was previously shared (Horton & Gerrig, 2005b). To measure how long it took Directors to initiate their descriptions, two time points were determined in the digital video-recording of the camera positioned behind the Director: (1) the first video frame in which the tangram on the current card became fully visible to the camera (since Directors were instructed not to turn a card until the Matcher had placed the previous card in the target area), and (2) the onset of the first content word that was part of the description of the current card[4]. These two time points were identified using the video annotating tool ELAN (Brugman & Russel, 2004) for descriptions in Phase 2[5].

3.3.9. Turns. Uninterrupted stretches of speech by a Director or Matcher were counted as turns. Instrumental actions performed by participants, such as Matchers

[4] Two triads were excluded from coding description onset time. For one triad, the video-recording of the camera behind the Director had been corrupted. For the other triad, the Director, despite instructions, previewed all cards multiple times before describing them.
[5] A third event was considered for some items as well: when the Matcher placed a card on the target area. I coded but excluded from analysis the description onset times for any cards that the Director turned over before the Matcher placed the previous card on the target area. In these cases, the Director, upon turning over her current card (despite instructions), may or may have not waited for the Matcher to finish placing her previous card on the target area; as such, the latency between time points (1) and (2) may not have reflected solely the planning of the current tangram's description.

placing cards on the target area were also counted as turns, even when these actions were performed in silence. If a Matcher ratified their acceptance of a Director's description in speech (e.g., by saying "*got it*", "*yeah*", or "*mm-hmm*") while simultaneously placing their card selection on the target area, their speech and instrumental action were considered to belong to the same turn. Exchanges at the beginning or end of a round that dealt with readiness to begin (e.g., D: *ready*? M: *yeah*) or concluding the round (e.g.: D: *and that's it*) were excluded so as not to inflate the turn count of the first and last items discussed. When non-verbal feedback, such as head nods or facial displays, were unaccompanied by speech they were counted as a turn but only if the conversational partner was attending it (e.g., if a Director nodded in response to their Matcher placing the card on the target area, the head nod was included as a turn only when the Matcher was looking at the Director, monitoring for their feedback).

3.3.10. Types of contributions. I coded the Matchers' responses that immediately followed the Director's initial description, or first turn. I used the same five categories described by Horton and Gerrig (2002): (1) *Acceptance*: the Matcher indicated successful identification through a simple "*okay*" or "*got it*", or by placing their selected card on the target area; (2) *Clarification request*: the Matcher requested clarification of some portion of the Director s previous description (e.g., asking "*is it like a perfect square*?" after hearing *"<um> this one is pretty solid / if you closed it up it would like just like a / large square with another square sitting on top / that's tilted or it could be a diamond <um> I don't know a big block / it doesn't have anything hanging off*"); (3) *Expansion request***:** the Matcher either implicitly requested an expansion of the previous description (often by saying "*umm*") or sought confirmation for a proposed expansion (e.g., asking ''*can also*

look like he's drinking something?'' after hearing ''*<uh> this looks like a person playing a trumpet* ?''); (4) *Interruption-acceptance*: the Matcher interrupted the Director's description to indicate acceptance; (5) *Interruption-query*: the Matcher interrupted the Director's description to ask for additional information or to indicate some kind of difficulty.

3.3.11. Errors. The number of errors involved in matching a target tangram was determined. There were three possible types of errors in this referential communication task, the last two being specific to Phase 1: (1) selecting the wrong tangram, (2) selecting a tangram instead of a joker card, and (3) selecting a joker card instead of a tangram.

3.3.12. Source monitoring. The source monitoring questionnaire results should be considered exploratory since Directors' responses, by design, were subject to interference: the questionnaires were administered at the end of the session, as opposed to at the end of Phase 1; encountering every item with both Matchers in Phase 2 could have interfered with Directors' explicit memory for how information had been shared in Phase 1. To deal with any response biases in reporting that items had been shared linguistically and physically in Phase 1 (in light of sharing all items linguistically and physically in Phase 2), I assessed d', which taps onto Directors' sensitivity for how information had been shared while taking response biases into account (Macmillan & Creelman, 1991). To calculate d', I classified each response as a hit, miss, false alarm, or correct rejection, with respect to the four conditions of co-presence. For example, if for a given tangram that had been shared only linguistically with a given Matcher in Phase 1 the Director said that it'd been shared linguistically and physically, their response was coded as a false alarm for the linguistically and physically co-presence condition, a miss for the

linguistically only condition, and a correct rejection for the physically only and new conditions.

3.3.13. Reliability. I coded all measures, except description onset and disfluencies, in their entirety. I chose to establish reliability for the somewhat subjective measures involving judgments about the segmentation and equivalency of units of meaning (i.e., idea units and reconceptualizations), as opposed to more objective measures that involved counting (e.g., words or hedges). To assess reliability for idea units and reconceptualizations, an undergraduate research assistant (blind to the Matcher's identity for each item in Phase 1) redundantly coded approximately 12.5% of the corpus for the number of idea units of tangram descriptions (total for Phase 1, and initial for Phase 2) and reconceptualizations (for items in Phase 2). This amounted to randomly choosing four triads, resulting in a total of 192 judgments for idea units and 128 judgments for reconceptualizations. We identified the same number of idea units for 78% of the cases, and for another 16% of the cases our idea units' count differed by only one idea unit. We also identified the same number of reconceptualizations for 74% of the cases, and for another 20% of the cases our reconceptualization counts differed by only one idea unit. For both measures our judgments were highly correlated: Pearson correlation $r = .98$, $p < .001$ for our idea units' count; Pearson correlation $r = .98$, $p < .001$ for our reconceptualizations' count.

Since the coding for description onset was done by me and three undergraduate research assistants, in order to assess reliability, we redundantly coded approximately 9%

of the corpus by randomly choosing three triads, resulting to a total of 91 measurements[6] of description onset. The mean absolute differences between my and the coders' measurements was 72.97 ms for the first coder, 81.51 ms for the second coder, and 107.31 ms for the third coder; these small differences arose primarily from disagreement about when the image on the Director's rotating card first became fully visible (time point 1) as opposed to disagreement about the beginning of the first content word (time point 2). My measurement for onset latencies was highly correlated with each of the other coders' measurement: with coder 1: Pearson correlation $r = .993$, $p < .001$; with coder 2: $r = .993$, $p < .001$; with coder 3: $r = .987$, $p < .001$. Critically, as revealed by a 4 X 4 ANOVA (4 coders x 4 co-presence conditions) there was no interaction between coder and co-presence condition (F (3, 87)= .76, *n.s.*).

[6] Five cases were excluded from analysis because the Director's description onset for the target card overlapped with the Matcher placing the previous card on the target area.

CHAPTER 4: Experiment 1 (Tangrams) Results and Discussion

4.1. Comparison of descriptions in Phase 1 for tangram items shared linguistically and physically versus only linguistically

Before examining whether the Directors' early linguistic choices in Phase 2 reflect sensitivity to how they shared items with their Matchers in Phase 1, I wished to establish whether Directors in Phase 1 discussed differently items that were shared both linguistically and physically with their Matchers compared to items that were shared only linguistically. I expected that pairs would have more difficulty with items that Directors had but Matchers didn't, since visual evidence can affect the process of grounding or establishing whether something is shared (Brennan, 2005, 1990; Clark & Krych, 2004; Gergle, Kraut & Fussell, 2004). Specifically, I expected that when Directors described tangrams that their Matchers didn't have, they would produce for the whole description more words, more idea units, use more hedging and metacomments, and would discuss the tangram over more turns. These expected patterns held, as Table 2 shows.

When Matchers didn't have the tangram being described, Directors' overall tangram descriptions had more words ($F1$ (1, 31)= 66.74, $p < .001$; $F2$ (1, 15)= 27.49, $p < .001$), more idea units ($F1$ (1, 31)= 62.74, $p < .001$; $F2$ (1, 15)= 34.04, $p < .001$), were more likely to contain hedges ($F1$ (1, 31)= 12.10, $p < .01$; $F2$ (1, 15)= 7.17, $p < .05$) and metacomments ($F1$ (1, 31)= 8.85, $p < .01$; $F2$ (1, 15)= 7.94, $p < .05$), and involved more turns ($F1$ (1, 31)= 26.75, $p < .001$; $F2$ (1, 15)= 10.28, $p < .01$).

Pairs also made significantly more errors when Directors described card that their Matchers didn't have ($F1$ (1, 31)= 16.48, $p < .001$; $F2$ (1, 15)= 16.16, $p < .01$): pairs made on average .20 errors per description (SD= .48) when discussing cards that

Matchers didn't have, while making only .05 errors (*SD*= .28) when discussing cards that Matchers had. Most of the errors that pairs made in Phase 1 (77% of the 62 errors) involved Matchers placing a tangram card in place of what should have been a joker card. The remaining errors involved cards that both the Directors and Matchers had: 15% of these errors involved Matchers placing the wrong tangram in place of a tangram they already had and only 8% of these errors involved Matchers placing a joker instead of a tangram they in fact had.

When Directors described tangrams that their Matchers had, the Matchers first contribution was an acceptance 50% of the time (*SD*= 50%), while when the Directors described cards that the Matchers didn't have it was an acceptance (involving putting up joker card) 38% of the time (*SD*= 48%). Conversely, when Matchers did not have the tangram, their first contribution was a clarification request 15% of the time (*SD*= 36%) and an expansion request 41% of the time (*SD*= 49%), while when they had it clarification and expansion requests dropped to 6% (*SD*= 24%) and 29% (*SD*= 45%), respectively. Contributions that involved Matchers interrupting their Directors to accept their description (Interruption: Acceptance) or to ask a question (Interruption: Query) totaled 15% of the cases where the Matcher had the tangram and 6% of the cases where they didn't. The distribution of acceptances versus clarification requests and expansion requests depended on whether Matchers had the card being described: for these contrasts, there was an interaction between contribution type and the condition of co-presence (acceptance vs. clarification request: $F1$ (1, 31)= 13. 81, $p < .01$, $F2$ (1, 15)= 7.62, $p < .05$; acceptance vs. expansion request: $F1$ (1, 31)= 9.00, $p < .01$; $F2$ (1, 15)= 6.82, $p < .05$).

Overall, although Directors shared tangrams linguistically with Matchers in both of these co-presence conditions, their cumulative descriptions and their interactions with their Matchers differed. When tangrams were shared only linguistically, Directors' total description of the card was elaborated, involving more words, idea units, hedges and metacomments. This stemmed largely from the fact that pairs discussed cards over more turns, during which Directors were more likely to have to respond to the Matcher's request for a clarification or expansion of their description or have to recover from an error. These findings are consistent with previous work showing that conversational partners adopt different grounding techniques depending on whether the speaker has visual evidence of the addressees understanding, with partners going through a lengthier process of checking that something has been mutually understood when relying only on spoken evidence (Brennan, 1990, 2005).

Thus, to the extent that the following sections of this chapter show evidence that Directors in Phase 2 elaborate their initial descriptions of tangrams that had been shared only linguistically in Phase 1 relative to tangrams that had been shared both linguistically and physically, this would suggest that the identity of the Matcher serves as a cue for the Director not only for whether the item was previously described but also for how the item had been grounded, depending on whether the Matcher previously had the item or not.

4.2. Definite expressions and meta-references

Having established how Directors discussed items that were shared both linguistically and physically versus only linguistically with their Matchers in Phase 1, I then turned to their early linguistic choices in Phase 2. I first asked whether their use of definite expressions and meta-references in their initial descriptions would reflect

whether a Matcher's identity served as a cue for shared information that varied probabilistically depending on the conditions of co-presence in Phase 1, or whether Matchers would instead serve as a binary cue for whether the information was mentioned versus not mentioned (or shared versus not shared) in Phase 1.

Directors produced a definite expression 31% of the time (*SD*= 46%) when a tangram had been shared linguistically and physically in Phase 1, 29% of the time (*SD*= 46%) when it had been shared only linguistically, 9% of the time (*SD*= 28%) when it had been shared physically, and 5% of the time (*SD*= 23%) when it was completely new to the Matcher in Phase 2. Figure 2 shows the means for the proportion of tangram items that included a definite expression or meta-reference in the Director's initial description, according to the conditions of co-presence in Phase 1 and Matcher identity.

The way information had been shared in Phase 1 affected the Directors' use of definite expressions and meta-references in their initial description in Phase 2: there was a significant main effect of the co-presence condition in Phase 1 ($F1$ (3, 93)= 26.57, $p <$ .001; $F2$ (3, 45)= 18.29, $p < .001$). Directors were more likely to use definite expressions and meta-references when they had shared an item linguistically and physically with a Matcher in Phase 1 than when the item was completely new ($F1$ (1, 31)= 40.63, $p < .001$; $F2$ (1, 15)= 26.22, $p < .001$). They were also more likely to use definite expressions and meta-references when they had shared an item only linguistically with a Matcher than when an item was completely new ($F1$ (1, 31)= 36.52, $p < .001$; $F2$ (1, 15)= 28.97, $p <$ .001) or when they had shared an item only physically ($F1$ (1, 31)= 25.38, $p < .001$; $F2$ (1, 15)= 20.67, $p < .001$). Directors did not distinguish through their definite expressions or meta-references between items shared both linguistically and physically and items

shared only linguistically (*F1* (1, 31)= .32, *n.s.*; *F2* (1, 15)= .14, *n.s.*), nor did they distinguish significantly between items shared only physically and completely new items (*F1* (1, 31)= 2.39, *p*= .13; *F2* (1, 15)= 1.30, *p*= .27).

There were no significant main effects of Matcher order or Matcher identity. The interaction between Matcher identity and Matcher order was, however, significant (*F1* (1, 31)= 6.91, $p < .05$; *F2* (1, 15)= 9.57, $p < .01$). This was because Directors, unexpectedly, tended to use marginally more definite expressions and meta-references in order ABBA than order ABAB with Matcher A (*F1* (1,31)= 4.12, *p*= .05; *F2* (1, 15)= 3.14, *p*= .10) and marginally more definite expressions and meta-references in order ABAB than order ABBA with Matcher B (*F1* (1,31)= 2.87, *p*= .10; *F2* (1, 15)= 3.08, *p*= .10).

Looking at definite expressions and meta-references on their own, it would seem that a Matcher's identity serves as a binary cue or constraint for how information had been shared in Phase 1: namely, indexing whether a tangram had been previously described or not. However, considering this finding together with the following measures provides a more nuanced understanding of what Directors remembered about how they had shared information with their partners.

4.3. Number of words

The number of words in the Directors' initial descriptions was taken as a proxy for the amount of propositional content they included. I asked whether adjustments in the number of words Directors used would reflect whether Matchers served as a cue for shared information that varied probabilistically depending on the conditions of co-presence in Phase 1, or whether Matchers would instead serve as a binary cue for whether

the information was mentioned versus not mentioned (or shared versus not shared) in Phase 1.

Directors used on average 17.49 words (*SD*= 11.04) in their initial description when a tangram had been shared linguistically and physically in Phase 1, 22.59 words (*SD*= 14.46) when a tangram had been shared only linguistically, 28.27 words (*SD*= 20.37) when a tangram had been shared only physically, and 23.15 words (*SD*= 16.12) when a tangram was completely new to the Matcher in Phase 2. Figure 3 illustrates the mean number of words in the Directors' initial descriptions, according to the conditions of co-presence in Phase 1 and Matcher identity.

The way information had been shared in Phase 1 affected the Directors' number of words in their initial description in Phase 2: there was a significant main effect of the co-presence condition of Phase 1 ($F1$ (3, 93)= 21.90, $p < .001$; $F2$ (3, 45)= 13.03, $p < .001$). Directors produced fewer words in their initial descriptions when a tangram had been shared linguistically and physically in Phase 1 than when it had been shared only linguistically ($F1$ (1, 31)= 17.52, $p < .001$; $F2$ (1, 15)= 17.09, $p < .01$), and in turn they produced fewer words when a tangram had been shared only linguistically in Phase 1 than when it had been shared only physically ($F1$ (1, 31)= 38.82, $p < .001$; $F2$ (1, 15)= 6.74, $p < .05$), demonstrating sensitivity to the degree that a tangram had been previously grounded. Surprisingly, Directors used more words when an item had been shared only physically in Phase 1 than when it was completely new in Phase 2 ($F1$ (1, 31)= 15.34, $p < .001$; $F2$ (1, 15)= 7.22, $p < .05$). Moreover, initial descriptions for completely new items did not differ significantly in their number of words from items that had been shared only linguistically in Phase 1 ($F1$ (1, 31)= .24, *n.s.*; $F2$ (1, 15)= .10, *n.s.*).

There were no significant main effects of Matcher order or Matcher identity. The interaction between Matcher identity and Matcher order was, however, significant (*F1* (1, 31)= 8.34, $p < .01$; *F2* (1, 15)= 17.39, $p < .01$), driven by Directors using more words with Matcher A in order ABAB than in order ABBA (*F1* (1, 31)= 7.34, $p < .05$; *F2* (1,15)= 4.14, $p = .06$). There was also a marginally significant interaction between Matcher identity and co-presence condition (*F1* (3, 93)= 2.56, $p = .06$; *F2* (3, 45)= 1.48, $p = .23$), because Directors tended to produced more words with Matcher B than with Matcher A for items that had been shared only physically in Phase 1 (*F1* (1, 31)= 2.81, $p = .10$, *F2* (1, 15)= 4.03, $p = .06$). This makes sense from a memory-based perspective since Matcher B was the most recent partner from Phase 1.

4.4. Idea units

The number of idea units that the Directors produced in their initial descriptions is a more accurate measure of assessing propositional content than word counts, since word counts could be inflated by hedging, metacomments, and self-interruptions. I asked whether adjustments in terms of the number of idea units Directors used in their initial descriptions would reflect whether Matchers served as a cue for shared information that varied probabilistically depending on the conditions of co-presence in Phase 1, or whether Matchers would instead serve as a binary cue for whether the information was mentioned versus not mentioned (or shared versus not shared) in Phase 1.

Directors' initial idea units are largely parallel to their initial word counts. Directors used on average 5.17 idea units (*SD*= 3.22) in their initial description when a tangram had been shared linguistically and physically in Phase 1, 6.54 idea units (*SD*= 4.03) when a tangram had been shared only linguistically, 7.85 idea units (*SD*= 5.58)

when a tangram had been shared only physically, and 6.80 idea units (*SD*= 4.76) when a tangram was completely new to the Matcher in Phase 2. Figure 4 illustrates the mean number of idea units in the Directors' initial descriptions, according to the conditions of co-presence in Phase 1 and Matcher identity.

Again, how information had been shared in Phase 1 affected the Directors' number of idea units in their initial description in Phase 2: there was a significant main effect of the co-presence condition of Phase 1 (*F1* (3, 93)= 15.72, $p < .001$; *F2* (3, 45)= 10.28, $p < .001$). Directors produced fewer idea units in their initial description when tangrams had been shared linguistically and physically in Phase 1 than when they had been shared only linguistically (*F1* (1, 31)= 13.71, $p < .01$; *F2* (1, 15)= 11.11, $p < .01$), and in turn, produced fewer idea units when tangrams had been shared only linguistically than when they had been shared only physically (*F1* (1, 31)= 20.59, $p < .001$; *F2* (1, 15)= 4.58, $p < .05$). As with word counts, Directors included more idea units in their initial descriptions of Phase 2 of tangrams that were shared only physically in Phase 1 than tangrams that were completely new (*F1* (1, 31)= 6.98, $p < .05$; *F2* (1, 15)= 4.70, $p < .05$). And as with word counts, Directors did not distinguish items that were shared only linguistically in Phase 1 from completely new items (*F1* (1, 31)= .48, *n.s.*; *F2* (1, 15)= .27, *n.s.*). As reflected by both word counts and idea units in initial descriptions of Phase 2, Directors attenuated items shared both linguistically and physically more so than items shared only linguistically, and attenuated items shared only linguistically more so than items shared only physically.

There were no significant main effects of Matcher order or Matcher identity. The interaction between Matcher identity and Matcher order was marginally significant (*F1*

(1, 31)= 2.73, p = .11; *F2* (1, 15)= 2.97, p= .11) because Directors tended to use more idea units with Matcher A in Matcher order ABAB than order ABBA (*F1* (1, 31)= 2.00, p = .17; *F2* (1, 15)= 1.28, p= .28). The interaction between Matcher identity and co-presence condition was also significant (*F1* (3, 93)= 4.49, $p < .01$; *F2* (3, 45)= 2.13, p= .11), because when describing items that had been shared only physically in Phase 1, Directors used more idea units in their initial descriptions to Matcher B than to Matcher A (*F1* (1, 31)= 5.85, $p < .05$; *F2* (1, 15)= 6.50, $p < .05$). As with the word counts, this pattern makes sense to the extent that Directors may remember better which tangrams had been shared only physically with Matcher B, since Matcher B was their most recent partner from Phase 1.

4.5. Reconceptualizations

To examine whether Directors would be more likely to depart from their original description of a tangram in Phase 1 depending on how they'd shared the tangram with their Matcher, I considered the number of idea units that were unique to Phase 2, defined as reconceptualizations. I asked whether Directors' use of reconceptualizations would reflect whether Matchers served as a cue for shared information that varied probabilistically depending on the conditions of co-presence in Phase 1 (such that there would be increasingly fewer reconceptualizations as the modalities through which the item was shared increased), or whether Matcher identities served as binary cues for whether the information was mentioned (or shared) before or not.

Directors introduced on average 2.02 reconceptualizations (*SD*= 2.23) in their initial description when a tangram had been shared linguistically and physically in Phase 1, 2.45 reconceptualizations (*SD*= 2.65) when a tangram had been shared only

linguistically, 3.41 reconceptualizations (*SD*= 4.01) when a tangram had been shared only physically, and 3.03 reconceptualizations (*SD*= 3.41) when a tangram was completely new to the Matcher in Phase 2. Figure 5 illustrates the mean number of reconceptualizations introduced in the Directors' initial descriptions, according to the conditions of co-presence in Phase 1 and Matcher identity.

Again, how information had been shared in Phase 1 affected the Directors' number of reconceptualizations in their initial description in Phase 2: there was a significant main effect of the co-presence condition of Phase 1 (*F1* (3, 93)= 9.48, $p < .001$; *F2* (3, 45)= 6.89, $p < .01$). When Directors described items that had been shared only physically in Phase 1, they introduced more reconceptualizations compared to items that were shared both linguistically and physically (*F1* (1, 31)= 21.44, $p < .001$; *F2* (1, 15)= 12.41, $p < .01$) or only linguistically (*F1* (1, 31)= 20.45, $p < .001$; *F2* (1, 15)= 7.29, $p < .05$). While Directors used more words and idea units in initial descriptions to items that had been shared only physically compared to completely new items, they did not distinguish these types of items significantly through their use of reconceptualizations (*F1* (1, 31)= 1.86, *p*= .18; *F2* (1, 15)= .89, *n.s.*). They also distinguished only marginally items that were shared only linguistically from items that were shared both linguistically and physically in Phase 1 (*F1* (1, 31)= 2.81, *p*= .10; *F2* (1, 15)= 2.68, *p*= .12). Notably, the pattern of reconceptualizations is strikingly similar to that for the counts of words and idea units in Directors' initial descriptions: when Directors had shared an item both linguistically and physically with their Matcher in Phase 1, they were less likely to elaborate their initial description or depart from their initial conceptualization compared to when they had shared an item only linguistically; and in turn, when they had shared an

item only linguistically they were less likely to elaborate their initial description or depart from their initial conceptualization compared to when they had shared an item only physically.

There were no significant main effects of Matcher order or Matcher identity. The interaction between Matcher identity and co-presence condition order was, however, significant by subjects (*F1* (3, 93)= 2.80, $p < .05$; *F2* (3, 45)= 1.71, $p = .18$), driven by the fact that Directors tended to produce more reconceptualizations for items that had been shared only physically with Matcher B than Matcher A (*F1* (1, 31)= 3.33, $p = .08$; *F2* (1, 15)= 3.81, $p = .07$). Again, as with the word counts and idea units, this pattern can be attributed to Directors remembering better which tangrams were shared only physically with Matcher B, since Matcher B was their most recent partner from Phase 1.

4.6. Hedges

To examine whether Directors adapted how provisionally they marked their descriptions depending on the conditions of co-presence in Phase 1, I considered their use of hedges and metacomments. First, I examined hedges to determine whether Directors would vary probabilistically their use of these markers of provisionality, depending on the conditions of co-presence (with less hedging as the modalities through which the item was shared increased), or whether their use would reflect a binary distinction between mentioned and unmentioned (or shared and not shared) items.

Directors used hedges in 12% (*SD*= 32%) of their initial descriptions of tangrams that had been shared linguistically and physically in Phase 1, in 18% (*SD*= 38%) of their initial descriptions of tangrams that had been shared only linguistically, in 29% (*SD*= 45%) of the descriptions of tangrams had had been shared only physically, and 23% (*SD*=

42%) of their initial descriptions of tangrams that were completely new to the Matcher in Phase 2. Figure 6 illustrates the mean proportion of tangram items with hedges in the Directors' initial descriptions, according to the conditions of co-presence in Phase 1 and Matcher identity.

Again, how information had been shared in Phase 1 affected whether Directors' included hedging in their initial description in Phase 2: there was a significant main effect of the co-presence condition of Phase 1 ($F1$ (3, 93)= 9.14, $p < .001$; $F2$ (3, 45)= 8.51, $p < .001$). When Directors described items that had been shared only physically in Phase 1, they were more likely to use hedges than when they described items were shared both linguistically and physically ($F1$ (1, 31)= 18.05, $p < .001$; $F2$ (1, 15)= 24.96, $p < .001$) or only linguistically ($F1$ (1, 31)= 11.60, $p < .01$; $F2$ (1, 15)= 9.84, $p < .01$). Through hedging, Directors marginally distinguished items that had been shared only physically from completely new items ($F1$ (1, 31)= 3.29, $p= .08$; $F2$ (1, 15)= 1.44, $p= .25$), as well as items that had been shared linguistically and physically from items that had been shared only linguistically ($F1$ (1, 31)= 2.60, $p= .12$; $F2$ (1, 15)= 4.21, $p= .06$). As with reconceptualizations, although some contrasts suggest more of a binary distinction between mentioned and non-mentioned items, there is a tendency for Directors to use an additional constraint regarding whether items had been shared physically or not. When the patterning of hedges is considered in conjunction with that of words, idea units and reconceptualizations, then the evidence that Directors distinguish the specific conditions of co-presence of Phase 1 becomes more compelling.

There was no significant main effect of Matcher identity, but there was a marginal main effect of Matcher order ($F1$ (1, 31)= 3.02, $p= .09$; $F2$ (1, 15)= 2.45, $p= .14$) as

Directors tended to produce more hedges in order ABAB. This was only because Directors happened to use more hedges with Matcher A in order ABAB than order ABBA ($F1$ (1, 31)= 8.72, $p < .001$; $F2$ (1, 15)= 12.00, $p < .01$), which also was reflected in an overall interaction between Matcher identity and Matcher order ($F1$ (1, 31)= 6.99, $p < .05$; $F2$ (1, 15)= 16.37, $p < .01$).

4.7. Metacomments

In addition to hedges, I considered metacomments as a way through which Directors could have marked their descriptions as provisional, since through metacomments they indicated their perspective and commitment to a description. I asked whether Directors would vary probabilistically their use of metacomments, depending on the conditions of co-presence (with fewer metacomments as the modalities through which the item was shared increased), or whether their use would reflect a binary distinction between mentioned and unmentioned (or shared and not shared) items.

Directors used metacomments in 6% (*SD*= 32%) of their initial descriptions of tangrams that had been shared linguistically and physically in Phase 1, in 12% (*SD*= 38%) of their initial descriptions of tangrams that had been shared only linguistically, in 14% (*SD*= 45%) of their initial descriptions of tangrams had had been shared only physically, and in 11% (*SD*= 42%) of their initial descriptions of tangrams that were completely new to the Matcher in Phase 2. Figure 7 illustrates the mean proportion of tangram items with metacomments in the Directors' initial descriptions, according to the conditions of co-presence in Phase 1 and Matcher identity.

Again, how information was shared in Phase 1 affected whether Directors' included metacomments in their initial description in Phase 2: there was a significant

main effect of the co-presence condition in Phase 1 (*F1* (3, 93)= 3.59, $p < .05$; *F2* (3, 45)= 4.07, $p < .05$). Directors were significantly less likely to include metacomments in their initial descriptions when a tangram had been shared both linguistically and physically with their Matcher in Phase 1 than all other conditions (versus only linguistically: *F1* (1, 31)= 7.75, $p < .01$, *F2* (1, 15)= 9.34, $p < .01$; versus only physically: *F1* (1, 31)= 7.52, $p < .05$, *F2* (1, 15)= 9.62, $p < .01$; versus new: *F1* (3, 93)= 5.16, $p < .05$; *F2* (1, 15)= 7.75, $p < .05$). They did not distinguish, through their use of metacomments, items that had been shared only physically from completely new items (*F1* (1, 31)= 1.00, *n.s.*; *F2* (1, 15)= 1.00, *n.s.*), or from items that had been shared only linguistically (*F1* (1, 31)= .41, *n.s.*; *F2* (1, 15)= .30, *n.s.*). But as Figure 7 suggests, the lack of difference between items that had been shared only linguistically versus only physically is due to the odd pattern of Directors using metacomments with Matcher A (but not Matcher B) very frequently. This pattern is due to Directors being significantly more likely to produce metacomments in their initial descriptions to Matcher A in order ABAB vs. order ABBA (*F1* (1,31)= 3.33, $p= .08$; *F2* (1, 15)= 9.62, $p < .01$). The effect was in fact driven by three Directors who, in order ABAB, used metacomments in 100% of their initial descriptions of items that had been previously shared only linguistically with Matcher A. This led to a marginally significant interaction between Matcher identity and co-presence condition (*F1* (3, 93)= 2.24, $p= .09$; *F2* (3, 45)= 1.57, $p= .21$), and a significant interaction by items between Matcher identity and Matcher order (*F1* (1, 31)= 2.90, $p= .10$; *F2* (1, 15)= 6.55, $p < .05$).

This odd pattern of Directors using metacomments very frequently for items that had been shared only linguistically with Matcher A in order ABAB obfuscates any

interpretation about whether Matchers in Phase 2 evoked how items were shared in Phase 1 in a probabilistic way or not. Perhaps because metacomments constituted a fairly heterogeneous set of expressions, including comments regarding the Director's perspective, certainty, or difficulty in describing the card ("*I don't know*", "*I guess*"), emotional reactions ("*oh my!*"), and explicit questions about the Matcher's knowledge of a domain, they may constitute less sensitive markers of provisionality than hedges. Nonetheless, through their use of metacomments, Directors still distinguished items that had been shared linguistically and physically with their Matchers in Phase 1 from all other items.

4.8. Disfluencies

The above measures have so far assessed how Directors marked definiteness (definite expressions and meta-references), provisionality (hedges and metacomments), and encoded propositional content (number of words, idea units, and reconceptualizations) in their initial descriptions. Since initial descriptions were defined as the Director's first turn, preceding the Matcher's contribution, it may be the case that some of the adjustments observed, particularly in terms of elaborating propositional content and marking descriptions as provisional, may have resulted from monitoring and responding to the Matcher's behavior as their initial description unfolded. Therefore, I also considered two measures that could tap into the Director's earliest planning: the disfluencies preceding the first content word and the amount of time taken to produce that first content word.

When considering the self-interruptions that Directors produced before the first content word of a description, there was some limited evidence that the conditions of co-

presence in Phase 1 mattered. Directors produced on average .09 self-interruptions (*SD*= .31) before the first content word of a description of a tangram that had been shared linguistically and physically in Phase 1, .15 self-interruptions (*SD*= .40) when a tangram had been shared only linguistically, .14 self-interruptions (SD= .37) when a tangram had been shared only physically, and .11 self-interruptions (*SD*= .32) for tangrams that were completely new for a Matcher in Phase 2. There was a marginal tendency for Directors to produce fewer self-interruptions preceding the description of tangrams that had been shared linguistically and physically compared to tangrams that had been shared only linguistically (*F1* (1, 31)= 2.42, *p*= .13; *F2* (1, 15)= 3.25, *p*= .09) or only physically (*F1* (1, 31)= 1.85, *p*= .18; *F2* (1, 15)= 2.65, *p*= .12)

The fillers that the Directors produced before the first content word of a description did not reflect sensitivity to how information was shared in Phase 1. Directors produced on average .43 fillers (*SD*= .58) before the first content word of a description of a tangram that had been shared linguistically and physically in Phase 1, .44 fillers (*SD*= .57) when a tangram had been shared only linguistically, .45 (*SD*= .61) when a tangram had been shared only physically, and .46 (*SD*= .57) for tangrams that were completely new for a Matcher in Phase 2. These small numerical differences were not significant.

Overall, disfluencies provided only a hint, specifically through self-interruptions, that Matchers served as a cue for how information had been shared in Phase 1, with Matchers cueing more effectively items that had been shared in two modalities over items that had been shared in only one modality.

4.9. Description onset time

Description onset time may be a more sensitive measure than disfluencies for whether the Directors' earliest planning reflects sensitivity for how information had been shared in Phase 1. On average, Directors took 2994 ms (*SD*= 1492 ms) to utter the first content word when describing tangrams that had been shared linguistically and physically in Phase 1, 3366 ms (*SD*= 1749 ms) for tangrams that had been shared only linguistically, 3467 ms (*SD*= 1940 ms) for tangrams that had been shared previously only physically, and 3225 ms (*SD*= 1682 ms) for tangrams that were new for the Matcher in Phase 2. Figure 8 shows the mean description onset latencies according to the co-presence conditions of Phase 1 and Matcher identity.

How information had been shared in Phase 1 affected how quickly Directors began their descriptions in Phase 2: there was a significant main effect of the co-presence condition of Phase 1 (*F1* (3, 63)= 3.67, $p < .05$; *F2* (3, 45)= 6.31, $p < .01$). Directors began describing tangrams that had been previously shared both linguistically and physically significantly more quickly than in all other conditions (versus only linguistically: *F1* (1, 21)[7]= 6.40, $p < .05$; *F2* (1, 15)= 12.13, $p < .01$); versus only physically: *F1* (1, 21)= 6.23, $p < .05$, *F2* (1, 15)= 15.62, $p < .01$); versus new: *F1* (1, 21)= 6.05, $p < .05$, *F2* (1, 15)= 3.53, $p = .08$). The numerical difference between tangrams that had been shared only linguistically vs. only physically was not significant (*F1* (1, 21)= .68, *n.s.*, *F2* (1, 15)= 1.74, $p = .21$), but there was a marginal tendency for Directors beginning to describe tangrams that had been shared only physically less promptly than

[7] The degrees of freedom in the by subjects' analyses reflect the fact that, although 30 Directors' description latencies were coded, eight were dropped from the by subjects' analysis because of missing data in one or more of the 16 cells (2 Matchers x 2 Matcher orders x 4 Co-Presence conditions) within subjects.

tangrams that were completely new to the Matcher (*F1* (1, 21)= 1.73, *p*= .20; *F2* (1, 15)= 4.84, $p < .05$). There was no main effect of Matcher identity (*F1* (1, 21)= .12, n.s.; *F2* (1, 15)= .15, *n.s.*), but there was a main effect of Matcher order (*F1* (1, 21)= 4.69, $p < .05$; *F2* (1, 15)= 4.90, $p < .05$) with Directors beginning their descriptions more promptly in order ABBA than order ABAB. This was driven by the fact that, in order ABBA, Directors with Matcher A (and not B) were for some reason faster to begin their tangram descriptions; the interaction between Matcher order and Matcher identity was also significant (*F1* (1, 21)= 16.23, $p < .01$; *F2* (1, 15)= 26.87, $p < .001$).

Overall, a Matcher's identity served as the most effective cue for Directors to begin their descriptions early when the information they were associated with had been shared both linguistically and physically.

4.10. Turns, Errors and Matcher's Contributions

In this section I leave behind audience design measures in order to consider how the conditions of co-presence shape co-ordination between partners. I ask whether, in addition to affecting the Directors' initial description and early processing, the conditions of co-presence may have also affected how Directors and Matchers coordinated until selecting the correct tangram, by examining the number of turns over which pairs discussed cards, the number of errors they made, and the Matchers' first contributions.

On average, pairs interacted over 2.99 turns (*SD*= 1.87) when they had previously shared a tangram linguistically and physically, 4.38 turns (*SD*= 3.99) when they had shared a tangram only linguistically, 4.69 turns (*SD*= 5.31) when they had shared a tangram only physically, and 4.26 turns (*SD*= 3.33) when the item was completely new for the Matcher in Phase 2. The way Directors and Matchers shared items in Phase 1

mattered; the co-presence condition of Phase 1 had a significant main effect (*F1* (3, 93)= 14.87, $p < .001$; *F2* (3, 45)= 5.86, $p < .01$). Pairs' interactions over items that had been shared linguistically and physically involved significantly fewer turns than all other conditions (versus only linguistically: *F1* (1, 31)= 36.50, $p < .001$, *F2* (1, 15)= 23.42, $p < .001$; versus only physically: *F1* (1, 31)= 27.75, $p < .001$, *F2* (1, 15)= 9.64, $p < .01$; versus new: *F1* (1, 31)= 40.05, $p < .001$, *F2* (1, 15)= 15.39, $p < .01$). Pairs' interactions over items that had been shared only linguistically, shared only physically, or were completely new for the Matcher did not differ significantly in their number of turns.

The way pairs shared items in Phase 1 also affected how many errors they made in Phase 2 (*F1* (3, 93)= 2.51, $p= .06$; *F2* (3, 45)= 2.90, $p < .05$). Pairs made a total of 32 errors in Phase 2 in the corpus (compared to 62 in Phase 1), all of which involved placing the wrong tangram on the target area. On average, pairs made .01 errors per tangram description (*SD*= .11) when they had shared a tangram linguistically and physically in Phase 1, .05 errors (*SD*= .26) when they had shared a tangram only linguistically, .04 errors (*SD*= .24) when they had shared a tangram only physically, and .03 errors (*SD*= .17) when the tangram was completely new for the Matcher in Phase 2. Pairs made significantly fewer errors when tangrams had been shared linguistically and physically relative to when they had been shared only linguistically (*F1* (1, 31)= 8.87, $p < .01$, *F2* (1, 15)= 7.35, $p < .05$) or only physically (*F1* (1, 31)= 4.23, $p < .05$; *F2* (1, 15)= 3.33, $p= .09$), and made marginally fewer errors relative to tangrams that were completely new to the Matcher in Phase 2 (*F1* (1, 31)= 2.95, $p= .10$, *F2* (1, 15)= 4.31, $p= .06$).

The proportion of clarification requests and interruptions (whether for an acceptance or a query) that Matchers offered in response to the Directors' initial

descriptions were comparable across the conditions of co-presence. However, there were differences in terms of how frequently Matchers responded with an acceptance (*F1* (3, 93)= 9.83, $p < .001$; *F2* (3, 45)= 7.15, $p < .05$) or a request for Directors to expand their description (*F1* (3, 93)= 9.04, $p < .001$; *F2* (3, 45)= 5.71, $p < .01$).

Matchers were significantly more likely to accept their Directors' initial description when they had shared the tangram linguistically and physically than in all other conditions (versus only linguistically: *F1* (1, 31)= 11.80, $p < .01$, *F2* (1, 15)= 16.58, $p < .01$; versus only physically: *F1* (1, 31)= 26.36, $p < .001$, *F2* (1, 15)= 17.02, $p < .01$; versus new: *F1* (1, 31)= 16.41, $p < .001$, *F2* (1, 15)= 10.19, $p < .01$). Matchers accepted their Directors initial description 77% (*SD*= 41%) of the time when a tangram had been shared linguistically and physically, 61% (*SD*= 49%) of the time when a tangram had been shared only linguistically, 58% of (*SD*= 50%) of the time when a tangram had been shared only physically and 59% (SD= 49%) of the time when it was completely new for them in Phase 2.

Conversely, Matchers were significantly less likely to request an expansion of the Directors' perspective when they had shared the tangram linguistically and physically than in all other conditions (versus only linguistically: *F1* (1, 31)= 11.22, $p < .01$, *F2* (1, 15)= 14.32, $p < .01$; versus only physically: *F1* (1, 31)= 21.10, $p < .001$, *F2* (1, 15)= 13.35, $p < .01$; versus new: *F1* (1, 31)= 19.20, $p < .001$, *F2* (1, 15)= 16.16, $p < .01$). Matchers requested an expansion of the Directors' perspective only 9% of the time (*SD*= 29%) when a tangram had been shared linguistically and physically, 21% of the time (*SD*= 41%) when it had been shared only linguistically, 23% of the time when it had been

shared only physically, and 24% of the time (*SD*= 42%) when it had been a completely new item for them in Phase 2.

Across these findings, a clear picture emerges: interactions over tangrams that had been shared linguistically and physically are distinguished from interactions over other tangrams. When Directions and Matchers both had a tangram in Phase 1, they interacted over fewer turns in Phase 2 and made fewer errors. Also, Matchers were more likely to accept, and less likely to request an expansion of, their Directors' initial description. The opportunity to fully ground a description in Phase 1—to ratify which was the exact referent of the Directors' description—was the most important factor accounting for the coordination between Directors and Matchers. Critically, although Directors did show sensitivity to the conditions of co-presence of Phase 1, as demonstrated by their initial partner-specific adjustments that I have reported in Sections 4.2. through 4.9., they weren't any more efficient at coordinating with their Matchers in Phase 2 over tangrams that had been shared only linguistically than over tangrams that had been shared only physically or were completely new.

4.11. Source monitoring

As suggested in Section 3.3.12, in their source monitoring questionnaires Directors showed a bias in responding that a tangram had been shared both linguistically and physically in Phase 1: while on average they identified correctly 86% of the tangrams that had been shared linguistically and physically, they also indicated incorrectly that 56% of the remaining tangrams had been shared linguistically and physically. The high proportion of false alarms for items that had been shared both linguistically and physically is consistent with the idea that Directors' responses about Phase 1 were subject

to interference from their experience in Phase 2, since questionnaires were administered at the end of the session, by which point Directors had already encountered every item linguistically and physically with both Matchers. Therefore, d' scores for each condition of co-presence were assessed instead.

Despite any interference from Phase 1, as Table 4 shows, there was some marginal evidence that d' scores differed across co-presence conditions (*F1* (3, 81)= 1.84, *p*= .15; *F2* (3, 45)= 1.78, *p*= .17). Directors were marginally more sensitive to identify items that had been shared linguistically and physically compared to items that had been shared only linguistically (*F1* (1, 27)= 5.76, $p < .05$; *F2* (1, 15)= 2.04, *p*=.17) or were completely new (*F1* (1, 27)= 3.58, *p*= .07; *F2* (1, 15)= 1.48, *n.s.*), at least according to the by subjects analyses. Directors did not exhibit significant differences in their sensitivity for items that had been shared only physically compared to items that had been shared linguistically and physically (*F1* (1, 27)= 1.23, *n.s.*; *F2* (1, 15)= .83, *n.s.*).

Previous work has failed to show evidence of partner-item associations in people's explicit recall. Horton (2007) demonstrated that in a non-communicative context that involved naming pictures of objects that had been previously associated with one of two partners, participants were faster to name the object associated with a particular partner when that partner was present. Nonetheless, this partner-specific facilitation in naming latencies was not significantly correlated with participants' explicit memory for the partner with whom items had been encoded. These findings were taken to suggest that explicit memory retrieval is not a primary mediator of the observed partner-specific facilitation and to reflect that partner-specific information becomes accessible on the basis of more implicit memory processes.

Although my findings here should be interpreted with caution, given that interference may have led to Directors' response biases, they do suggest that explicit memory retrieval should not be ruled out of the process of audience design. In this experiment, involving a task with a *communicative* context where participants named pictures as part of a joint activity with joint goals, the associations between partners and shared information seem to be reflected in the participants' explicit responses. This is in line with findings that partner-specific effects can be eliminated when the situation is not interactive, which have been taken to suggest that when people don't have the opportunity to ground information with a partner they may form weaker partner-specific associations (Brown-Schmidt, 2009a). Indeed, in my interactive task, the Directors' use of definite expressions and meta-references (including expressions like "*it's already come up before*", "*from the last time*", and "*again*") points out that implicit memory processes cannot fully account for the assessment of common ground.

4.12. Summary

To summarize, overall, Directors' initial descriptions of tangrams in Phase 2 reflected sensitivity to how they had shared them with their Matchers in Phase 1. Their adjustments were appropriate to how tangrams had been previously grounded in Phase 1. For instance, Directors were more likely to use definite expressions and meta-references for items they had described to their Matchers in Phase 1 (whether Matchers had these items or not) than for items they hadn't described to their Matchers. That is, when encountering an item they had previously described to their Matchers, Directors were more likely to mark a description as referring to a shared item. At the same time, Figures 3 through 7 illustrate a consistent pattern that suggests that Matchers didn't just cue a

binary distinction about whether the item had been described before or not, but rather cued more specifically how the item had been previously shared. For instance, the Directors' number of words and idea units, and to a lesser extent their hedges and reconceptualizations, distinguished whether items described in Phase 1 had been fully grounded or not: descriptions of tangrams that had been shared only linguistically in Phase 1 were more elaborated in terms of their words and idea units, were marked as more provisional and were more likely to involve a reconceptualization compared to those of tangrams that had been shared both linguistically and physically. In fact, Directors described tangrams that had been shared only linguistically with as many words and idea units as they described tangrams that were completely new for their Matchers in Phase 2.

The adjustments Directors made when distinguishing completely new items from items that had been shared linguistically and physically are in line with findings from earlier studies demonstrating new/old distinctions. Compared to new information, information that is old for the partner is attenuated in terms of words and idea units (Galati & Brennan, 2010; Horton & Gerrig, 2002), involves less hedging (Horton & Gerrig, 2002; Brennan & Clark, 1996) and fewer reconceptionalizations (Horton & Gerrig, 2002, 2005b), and is likely to be marked with definite expressions (Brennan & Clark, 1996); these findings are all consistent with how new items were distinguished from items that had been shared linguistically and physically in this experiment. Moreover, while Horton & Gerrig (2005b) failed to show differences in description onset latencies between old vs. new items in a referential communication task where Directors could use their Matcher as a cue for what categories of information they had previously

shared, I have found evidence that Directors did distinguish between old (fully shared) and completely new items, even with Matchers serving as a cue for shared information on an item-by-item basis. Because for my coding I used both video (to determine when Directors first saw the current tangram) and audio (to determine when they produced the first content word relevant to the description), I likely obtained more accurate description onset latencies than Horton & Gerrig (2005b), who used only audio and began measuring the onset of planning a description after the Matcher said "okay" in response to matching the previous card.

More important, this study provides what may be the first evidence that beyond distinguishing new from old, fully shared items, people distinguish how they had shared information with their conversational partners. Descriptions of items that had been shared only linguistically fell somewhere in between those of completely new items and items that had been shared linguistically and physically: they had some characteristics of descriptions of items that had been fully grounded in Phase 1 (comparable in terms of definite expressions) but also of descriptions of items that were completely new to the Matcher (being elaborated in terms of words and idea units).

A consistent and somewhat surprising finding in this study was that, in terms of word counts and idea units, Directors elaborated descriptions of items that had been shared only physically more than descriptions of items that were new for the Matcher in Phase 2. One possible explanation is that this effect is an artifact of the current design, arising from the fact that the conditions of co-presence were complementary across Matchers in Phase 1 (namely, items that had been shared only linguistically with one partner had been shared only physically with the other, and items that had been shared

both linguistically and physically with one partner had not been shared at all with the other partner). Therefore, items shared only physically may have, for some reason, been more distinctive than new items simply on account of being shared only linguistically with the other partner. Another possible explanation is that, to the extent that a Matcher's identity readily cues what happened in Phase 1, Directors accessed an actual episodic memory trace when determining that an item had been shared only physically, whereas they determined that an item had not been shared at all by inference—by failing to access an episodic trace for that partner-item association. Determining that an item had been shared but not described in Phase 1 through accessing an episodic trace may have led Directors to elaborate their initial descriptions more pronouncedly. This is supported by findings of an effect of Matcher identity or an interaction between Matcher identity and Matcher order for some measures: when items had been shared only physically, Directors used more words and idea units with Matcher B than Matcher A, and tended to produce more reconceptualizations with Matcher B. Since Matcher B was always the most recent partner from Phase 1, a stronger memory for the items that were shared only physically with Matcher B can explain why these items involved the most pronounced elaboration. The Directors' explicit source memory also lends support to this idea: Directors showed greater sensitivity at recalling explicitly that an item that had been shared only physically compared to recalling that an item was completely new.

That Directors in my study made distinctions in their initial descriptions according to the conditions of co-presence in Phase 1 demonstrates that a partner's identity can successfully cue the status of shared information on an item-by-item basis. While in previous studies conversational partners served as more global cues for shared

information—for instance, cuing category associations (e.g., with A, I have matched cards of dogs, not of turtles) or the information status of a more extended stretch of discourse (e.g., A has heard this entire story before), speakers here were able to distinguish the information status of items even when not able to group them into categories (since they were all tangrams) or into a more extended, coherent discourse structure that linked them together (since in Phase 1 items shared linguistically and physically were interspersed with items shared only linguistically).

The fact that some measures exhibit a binary distinction between items that had been described versus not (e.g., in terms of definite expressions) or between fully grounded items versus other items (as reflected by description onset latencies and to a lesser extent self-interruptions) does not imply that Directors did not encode how the information was shared in their episodic memory traces. Rather than reflecting a binary representation of what is indexed in episodic traces for shared information, these findings suggest instead that Directors chose an appropriate adjustment in response to their Matchers' informational needs: when Directors were describing an item they had already described before, even if their Matcher did not have it, they may have reasoned that by marking the description explicitly as shared by using a definite expression, they would facilitate their Matchers in identifying the card. When these findings are considered in conjunction with the adjustments Directors made in terms of their word counts, idea units, reconceptualizations, and hedges, which exhibit sensitivity to the specific conditions of co-presence in Phase 1, the warranted conclusion is that Matchers do in fact cue how information was shared in Phase 1. What Directors do is in fact appropriate in light of how information has been previously grounded: on one hand they mark items

shared only linguistically as previously mentioned, and on the other hand they include more detail in their descriptions, and make their descriptions more provisional.

In this light, consider the description onset findings: Directors in this experiment were faster to begin describing items that had been previously shared both linguistically and physically relative to all other items. The description onset latencies here may be interpreted as Directors not differentiating between items that had been shared only linguistically, only physically or new, but an alternative interpretation is that these three conditions raise for Directors different sorts of considerations for how to most appropriately package the information for their Matchers, including decisions about whether to mark definiteness (as they do with items shared only linguistically) or whether to depart from the terms of an earlier descriptions (as they do with items shared only physically and new items, and to a lesser extent items shared only linguistically). Describing a partially shared item is not necessarily easier or more straightforward than describing a completely new item, and these considerations during early utterance planning may have caused delays in description onset in all three conditions, relative to the more straightforward case of fully shared items. In other words, while description onset is typically taken to reflect the ease (or difficulty) of assessing shared information, here, it may also reflect the ease (or difficulty) of selecting an appropriate grounding strategy for utterance planning. For instance, longer description onsets may reflect the fact that speakers are evaluating whether a partially shared item warrants using a definite expression or reconceptualizing the perspective of an earlier description without increasing their conversational partner's processing cost. In order to understand what people remember about their prior experiences with conversational partners, it is

important to consider behavioral adjustments at different grains of linguistic processing to gain a holistic picture for both early processing and the early strategies that people adopt.

Finally, it should be noted that even though Directors made early, partner-specific adjustments in response to the informational needs of their Matchers, the Matchers did not always benefit from these adjustments. For instance, although Directors marked items shared only linguistically as mentioned by increasing their use of definite expressions and decreasing their use of reconceptualizations and hedges, pairs interacted no more efficiently when discussing these items compared to items that had been shared only physically or that were completely new for the Matcher. In all of these conditions they interacted over more turns, made more errors, and Matchers were more likely to request an expansion of the description than when an item had been shared both linguistically and physically. In other words, although Directors distinguished in their initial descriptions how they had shared items with their Matchers in Phase 1, the efficiency of their interactions with their Matchers was determined by whether items in Phase 1 had been fully grounded.

This suggests that the factors guiding audience design and the factors that affect coordination between conversational partners are not always identical. Although a particular behavioral adjustment may be potentially informative and may be mediated by the speakers' intentions toward the addressee, it may not be processed effectively by the addressee. These three criteria –the behavior being informative, mediated by the speakers' intentions, and recognized by the addressee—are necessary to make a particular behavioral adjustment communicative (as discussed by Brennan & Williams, 1995). In

other words, some of the Directors' markers for shared information (e.g., marking items shared only linguistically with definite expressions) were not always communicative, despite the Directors' intentions, insofar as these cues were generally not helpful for Matchers and did not facilitate the coordination of the pair. In this respect, the factors affecting audience design and coordination can be dissociated. Critically, although measures assessing the efficiency of coordination may not reveal sensitivity to how information had been previously shared between conversational partners, measures of audience design do. Addressees can cue speakers for how information had been previously shared, even if they don't always benefit from the linguistic cues that speakers intentionally provide for them.

CHAPTER 5: Experiment 2 (Category Exemplars) Materials and Coding

While in Experiment 1 I examined whether the conditions of co-presence in Phase 1 would affect partner-specific adjustments in Phase 2 at the level of utterance planning, in Experiment 2 I aimed to examine whether they would affect partner-specific adjustments at the level of articulation. Instead of using tangrams, which lack conventional labels and are typically difficult to describe, in Experiment 2 I used line drawings of common objects with the intention of culling lexically identical expressions, as these materials are easily lexicalized and typically involve using the same referring expressions across rounds. I then assessed, through a separate experiment (Experiment 3), whether for lexically identical expressions culled from Experiment 2 intelligibility in Phase 2 depended on the conditions of co-presence in Phase 1. Recall that the participants in Experiment 1 (described in Section 2.2.) also took part in Experiment 2, and that the order of Experiments 1 and 2 was counterbalanced across triads of participants.

In addition to excising lexically identical expressions for subsequent experiments, in Experiment 2 I made some exploratory assessments regarding the ease with which Directors access shared information. Since the items for this experiment are generally lexicalized easily, I did not necessarily anticipate that description onset would exhibit partner-specific adjustments sensitive to the co-presence conditions of Phase 1. Instead, I examined the disfluencies preceding the first content word that was part of the description of a given card. Moreover, I assessed the Directors' explicit memory for how they had shared items with each of their Matcher in Phase 1 based on their responses in the source monitoring questionnaire.

In this chapter I describe the materials used in Experiment 2 and detail the method involved in examining early disfluencies and source monitoring; I report these results in Chapter 6. In Chapter 7, I introduce Experiment 3 and describe the criteria I used to determine which sets of words would be eligible for inclusion, and how I sampled and excised the final set of stimuli for that experiment; I report the results of that experiment in Chapter 8.

5.1. Materials

5.1.1. Stimuli. Two identical sets of 16 line drawings of common objects from different categories, to be used by the Director and Matchers, were printed on 3 x 5 inch index cards. The category exemplars were grouped into two sets of eight cards; each set included an item from the following eight categories: animal, vehicle, food, clothing, instrument, furniture, body part, and tool. Directors and Matchers used each of these sets when playing for two separate sets of four rounds. Two additional items from different categories constituted the Matchers' ninth card in Phase 2, for each of the two sets. Line drawings for these common objects were obtained from the database of the International Picture Naming Project, which includes line drawings from the Snodgrass and Vanderward (1980) set and other sources. The 16 target items were selected to have their most common label be multisyllabic, and according to norming studies to have similar visual complexity (Szekely et al., 2003) and frequency, as measured by their log HAL frequency (Balota et. al, 2007). The 16 target items are listed in Appendix D, along with their category, visual complexity score and log HAL frequency. Two additional images of jokers were printed on 3 x 5 inch index cards to be used by Matchers in Phase 1.

5.1.2. Source monitoring questionnaire. As described in Section 3.3.12, two of the four pages of the source monitoring questionnaire were for the sets of category exemplar items. The order of the pages corresponded to the order in which participants had played with these sets of cards in their session. The pages for the category exemplar sets listed pictures for each of the 16 target category exemplars. The Directors' questionnaire included two columns next to each picture, one with the heading that asked them how they experienced each card with Matcher A in Phase 1 (when playing with the 6-slot board), and a second with the corresponding heading for Matcher B. For each picture, Directors had to chose one of the following options with respect to Matcher A: (a) both I and Matcher A had it, (b) only I had it (Matcher A used a joker), (c) only Matcher A had it (I placed it in slot 5 or 6), (d) neither I nor Matcher A had it, (e) I don't remember. The corresponding options with respect to Matcher B were listed for each item in the second column. As pointed out earlier, Matchers were also given an analogous questionnaire, but only the Directors' responses are of interest here.

5.2. Transcription

For each triad all four rounds of both sets of category exemplars were transcribed in detail, including contributions by both the Matcher and the Director. The same conventions as for transcribing the tangrams trials in Experiment 1 were used: transcripts included annotations of fillers such as "uh" or "um", pauses, lengthening of vowels, interruptions (both self-interruptions and interruptions by the Matcher) and restarts. Instrumental actions performed by the participants (such as the Matcher placing a card on the target area) and non-verbal feedback, such as head nods and facial displays, were also annotated in the transcripts.

5.3. Coding

5.3.1. Disfluencies. As with Experiment 1, disfluencies were assessed because they can signal the information status of a referring expression and may exhibit sensitivity to how information had been shared with a Matcher. Again, the total number of disfluencies (fillers and self-interruptions) preceding the first content word that was part of the description of the current card was determined. Typically, the first content word was the noun describing the target referent, like *motorcycle* or *spider*. Disfluencies preceding the first content word were identified for descriptions in Phase 2.

5.3.2. Source monitoring. I classified each of the Directors' responses in the source monitoring questionnaire as a hit, miss, false alarm, or correct rejection with respect to each of the four co-presence conditions (linguistically and physically co-present, only linguistically co-present, only physically co-present, and completely new).

CHAPTER 6: Experiment 2 (Category Exemplars) Results and Discussion

6.1. Disfluencies

To tap into the Directors' earliest planning, I considered the disfluencies—fillers and self-interruptions—preceding the first content word. Directors produced fillers before the first content word of a description in Phase 2 in only 7.5% of the trials. The distribution of fillers did not reflect sensitivity to how information had been shared in Phase 1 (*F1* (3, 93)= 1.36, *n.s*., *F2* (3, 45)= .56, *n.s*.). Directors produced on average .08 fillers (*SD*= .28) before the first content word of a description of an item that had been shared linguistically and physically in Phase 1, .07 fillers (*SD*= 26) when an item had been shared only linguistically, .05 (*SD*= .23) when an item had been shared only physically, and .09 (*SD*= .29) for items that were completely new for a Matcher in Phase 2. The difference between completely new items and items that had been shared only physically was significant by subjects (*F1* (1, 31)= 5.74, $p < .05$; *F2* (1, 15)= 1.34, $p = .26$). When describing category exemplars, Directors produced only three self-interruptions in the entire corpus, so these were not considered further.

Overall, disfluencies weren't informative about how Directors had shared items in Phase 1 with their Matchers: at this juncture, it is unclear whether these null effects reflect a lack of distinctive influence of the conditions of co-presence of Phase 1, or merely the fact that, for Directors, describing these items was too easy for any partner-specific effects on utterance planning to be manifested in terms of disfluencies. Directors' explicit memory reports can elucidate this issue.

6.2. Source monitoring

As with Experiment 1's tangram items, Directors' responses in the questionnaire for Experiment 2's category exemplar items showed the same bias for items that had been shared both linguistically and physically in Phase 1: while they identified correctly on average 79% of the items that had been shared linguistically and physically, they indicated falsely that another 54% of the remaining items had been shared linguistically and physically as well. These estimates were remarkably similar to those for the tangrams in Experiment 1 (86% and 57%, respectively), suggesting that the tendency to judge items as shared linguistically and physically did not depend on the complexity of the item.

There was evidence that sensitivity to the conditions of co-presence differed (*F1* (3, 81)= 5.54, $p < .01$; *F2* (3, 45)= 2.63, $p= .06$), despite any interference from Phase 2, as illustrated as illustrated in the pattern of d' scores in Table 4. Directors were more sensitivity to identify items that had been shared linguistically and physically compared to items that had been shared only linguistically (*F1* (1, 27)= 10.81, $p < .01$; *F2* (1, 15)= 5.23, $p < .05$) or items that had been completely new (*F1* (1, 27)= 5.97, $p < .05$; *F2* (1, 15)= 1.85, *n.s.*); their sensitivity did not differ significantly compared to items that had been shared only physically (*F1* (1, 27)= .49, *n.s.*; *F2* (1, 15)= 1.19, *n.s.*). Items that had been shared only physically also demonstrated higher sensitivity scores compared to items shared only linguistically (*F1* (1, 27)= 9.99, $p < .01$; *F2* (1, 15)= 8.46, $p < .05$) or items that were completely new (*F1* (1, 27)= 4.56, $p <. 05$; *F2* (1, 15)= 2.63, $p= .13$).

6.3. Summary

In terms of their use of fillers and self-interruptions, Directors did not distinguish significantly how they had previously shared items with their Matchers in Phase 1. But critically this was only because the items were easy to name, and not because Directors did not remember how they had shared them previously with their Matchers.

Indeed, as with Experiment 1, Directors distinguished how items had been shared, as reflected by their responses in the source monitoring questionnaire. As with Experiment 1, responses on the source monitoring questionnaire were likely subject to interference, and should be interpreted cautiously. Although the level of effort required by conversational partners to arrive at the mutual belief that they had understood each other was much lower with category exemplars than with tangrams, Directors demonstrated that, despite these differences in grounding criteria between the two experiments, they still indexed how information had been shared.

A consideration here is that in Experiment 2, given the more conventionalized and less variable labels of referents, Directors generally did not have to negotiate items that Matchers didn't have any more than those that Matchers did have. Unlike Experiment 1, Matchers in Experiment 2 were typically quick to confirm that they didn't have a common object (e.g., a pineapple) and place a joker in its place. Since in Experiment 2 interactions over described items in Phase 1 were similar, regardless of whether the Matcher had the described item or not, the experience of sharing an item linguistically and physically was comparable to the sum of the experience of sharing an item linguistically only and physically only. On the other hand, in Experiment 1, the experience of sharing an item linguistically and physically was not equal to the sum of

the experience of sharing an item linguistically only and physically only. Recall that in Phase 1 of Experiment 1, pairs used different grounding techniques depending on whether the Matcher had the card or not: pairs' interactions over items shared only linguistically were longer, lasting over more turns, with the Director having to respond to more requests for an expansion or clarification compared to items shared both linguistically and physically (see Section 4.1). Therefore, episodic traces resulting from sharing items linguistically and physically in Phase 1 of Experiment 1 would not be comparable to the sum of sharing items only linguistically and only physically. In light of this, the partner-specific effects observed, particularly in Experiment 1, may be best framed as a function of the conditions of co-presence or the affordances of the communicative situation rather than as a function of multimodality. This impacts the framing of my results, but not their implications. That Directors' descriptions in Phase 2 demonstrate sensitivity to how information was shared in Phase 1, in both Experiments 1 and 2, as indicated by their appropriate adjustments or explicit memory, still offers insight into their underlying memory representations for shared information.

Of course, conversational partners in Experiment 2 did sometimes negotiate their referring expressions for a category exemplar. Consider the following example from triad 26, in which the Director in Phase 1 describes the trumpet card to Matcher B:

D26: <um> / musical instrument /
I<I> / don't know what this is called /
it's<s>* /
it's a musical instrument
<um> not a flute /
clarinet ? /
maybe that's what it's called ? /
M_B: is it long ?
D26: <um> / it's short / it's *almost* like* /
damn

what is it called ? %laugh
<um> / it's not that long /
it's *kind of* short
<um> /

M_B: does it have buttons ?

D26: yeah
it has three buttons

M_B: [*puts up trumpet card*]

D26: yeah

Remarkably, even though Matcher B does have the target card and it is in fact the only musical instrument among the cards in her staging area, she still negotiates with her Director what the referent might be until she has sufficient evidence for current purposes that she has the right card[8].

Later, in Phase 2 (Round 3), when the Director describes the trumpet card to the same Matcher, she demonstrates clear evidence that she recalls their prior interaction, as indicated both by her use of a definite expression and by laughing:

D26: %laugh the musical instrument %laugh /

M_B: %laugh
[*puts up trumpet card*]

Compare this exchange with how the Director then describes the trumpet card to Matcher A (Round 4). She no longer uses a definite expression and provides additional detail about the item (in fact, the critical detail that in Phase 1 had helped Matcher B identify the card) in a complementizer clause:

D: a musical instrument that has three buttons /

M_A: a trumpet ?

D: yeah

M_A: %laugh
[*puts up trumpet card*]

[8] Since participants were not explicitly told that in Experiment 2 items would be from different categories, the Matcher may still have wished to rule out the possibility that the Director was describing another musical instrument that she didn't have.

These sorts of multi-turn exchanges in Phase 1 were not typical in Experiment 2, but occurred occasionally, either when Directors had difficulty finding the right referring expression for the item as with the example above or when Matchers requested an expansion of the Directors' description as with the example below:

D27: a<a> spider in its web /
M_B: and it has dots on the bottom ? /
D27: yes sir /
M_B: okay /
[*puts up spider card*]

When the distribution of the definite expressions in Experiment 2 was considered (*n*=15), two thirds of them occurred for referents that had been previously described: 8 of the definite expression were for items that had been shared linguistically and physically, 2 were for items that had been shared only linguistically, and 2 were for items that had been shared only physically.

Overall, although Directors did not demonstrate reliable differences across co-presence conditions in their distribution of fillers, probably because these items were so easily lexicalized, their explicit recall did show evidence of differentiating how items had been shared. Other aspects of encoding propositional content, such as Directors' use of definite expressions, also offer some corroborating evidence that Directors tracked how items had been shared. Experiment 3 investigates whether this distinction is reflected in partner-specific adjustments in the Directors' intelligibility of referring expressions as well.

CHAPTER 7: Experiment 3 (Rating Clarity) Method

Lexically identical expressions from Experiment 2 were culled in order to have their intelligibility assessed in Experiment 3. Earlier work has shown that expressions repeated to an old partner are less intelligible than those repeated to a new partner (Galati & Brennan, 2010). I wanted to extend these findings to determine whether the conditions of co-presence would affect intelligibility as well. Specifically, I asked if referring expressions in Phase 2 would be less intelligible if the items had been shared through both modalities in Phase 1, as compared to only one modality, or not having been shared at all.

7.1. Participants

Forty-one Stony Brook University students participated in exchange for credit toward a research requirement in a psychology course. The data of one participant were excluded because she did not follow instructions. Of the remaining 40 participants, 26 were female, 14 male. All participants were native speakers of English and reported no hearing problems. None had participated in Experiments 1 and 2.

7.2. Materials

The items that listeners heard in Experiment 3 were selected from Experiment 2 according to the following selection criteria and selection process.

7.2.1. Selection of eligible tokens. With the objective of sampling triplets of items (one from Phase 1, and two from Rounds 3 and 4 in Phase 2) to use in this experiment, I used the following five criteria (also used by Galati and Brennan, 2010, except for the second criterion) to determine eligible triplets for each of the 16 target category exemplar referents:

1) The first mentions of the expression in each of the three rounds (Round 1 or 2, Round 3, and Round 4) had to be lexically and syntactically identical. For instance, I did not consider the three tokens of "*scissors*" in the triplet "*a pair of scissors*"-"*scissors*"- "*scissors*" to be lexically identically because in the first mention it was not the head of the noun phrase.

2) Expressions had to have the same phonetic environment, that is, be preceded and followed by the same phonemes.

3) Expressions had to have the same pause environment. The pause environment was considered to be the same when all tokens of a triplet were (a) followed but not preceded by a pause, (b) preceded but not followed by a pause, (c) both preceded and followed by a pause (e.g., when Directors said simply "*scissors*"), or (d) neither preceded nor followed by a pause. This ensured that the expressions were in all in the same position in the utterance (initial, medial, or final), since speakers tend to elongate words at the end of an utterance (Cooper & Paccia-Cooper, 1980).

4) If any of the tokens of a triplet had list-final, falling intonation, which sometimes Directors used when describing the final, eighth item in Phase 2, or the fourth item in Phase 1, the triplet was excluded.

5) Finally, I excluded any triplets for which any of the tokens contained laughing, coughing, a self-interruption, or overlapping speech from the Matcher.

A total of 134 triplets fulfilling these criteria were contributed by 25 Directors. The eligibility of triplets was not assessed for three Directors for whom the high quality audio recorder either failed to record or the recording was partially lost, and one Director who did not sign a release form to allow use of his data in subsequent experiments.

7.2.2. Sampling process. I aimed to represent each of the 16 items in all four pairings of co-presence conditions for the two Matchers (shared linguistically and physically with Matcher A and not shared at all with Matcher B in Phase 1; shared linguistically with Matcher A and physically with Matcher B in Phase 1; shared physically with Matcher A and linguistically with Matcher B in Phase 1; and not shared at all with Matcher A and shared linguistically and physically with Matcher B in Phase 1), and in both Matching orders (ABAB and ABBA). My approach involved sampling from the maximum number of contributing Directors, while balancing the number of contributions across the four pairings of co-presence conditions and the two Matcher orders.

The resulting sample consisted of 84 triplets from 25 contributing Directors. Table 3 shows the distribution of the sample's triplets across co-presence conditions and Matcher orders.

Words were excised to include their complete onsets and offsets; this was done by examining their waveforms and listening to the edited results. Stimuli were processed through a digital to analog converter (12 bit; 16 kHz rate); for a few triplets from Directors who spoke either too softly or too loudly the amplitude was adjusted in Goldwave. Crucially, the same adjustment in amplitude was applied to all three tokens within each triplet to avoid affecting their relative intelligibility.

7.3. Procedure

Listeners accessed the 84 triplets in one of two semi-randomized orders (the second order being the reversed order of the first); they heard all three versions of each expression. They were instructed to rate on an answer sheet how clear the three tokens of each triplet were with respect to one another. They accessed the items on the desktop of a computer in 84 folders, labeled Trial 1 through 84, each of which contained audiofiles with the three versions of a given expression; the audio files were coded with random fish or bird names (e.g., *egret*, *flamingo*, *heron*). Listeners were instructed to open all three audiofiles, and play them as many times as necessary in order to assess the relative clarity of the three words with respect to one another. They rated the three tokens of a triplet for clarity on a single scale from 1-5, where 1 was low and 5 was high clarity. If two or even all three audiofiles sounded equally clear, they could be assigned the same rating. After completing the experiment, participants were debriefed.

7.4. Design and Analyses

Two sets of ANOVAs were done on the relative clarity scores: one involved a one-way ANOVA comparing the information status of mentions in Phases 1 and 2 (first mention (Phase 1) vs. Phase 2 mentions for items that had been shared linguistically and physically, shared only linguistically, shared only physically, or completely new); the other involved a 2 x 2 x 4 ANOVA with Matcher identity (Matcher A vs. Matcher B), Matcher order (ABAB vs. ABBA), and co-presence condition (linguistically and physically shared, linguistically shared, physically shared, completely new) as within

subjects factors[9]. Effects of item order are not reported, as preliminary analyses revealed that they were not significant.

[9] For the ANOVAs in Experiment 3, I considered items to be the triplets of excised referring expressions (*n*= 84) as opposed the referents (*n*= 16), seeing that not all referents were represented in the sample in all 16 conditions (2 Matcher Identity x 2 Matcher Order x 4 Co-Presence), as pointed out in Section 7.2.2. With one or more of the 16 cells for half of the referents missing, an ANOVA by-referents could not be conducted.

CHAPTER 8: Experiment 3 (Rating Clarity) Results and Discussion

Listeners' clarity ratings reflected differences according to how Directors in Experiment 2 had shared items with their Matchers. The mean clarity rating for the first mention of a referring expression in Phase 1 was 3.75 (*SD*= 1.12), while for repetitions of the referring expression in Phase 2, the mean clarity rating was 3.48 (*SD*= 1.17) when the item had been previously shared both linguistically and physically, 3.56 (*SD*= 1.22) when the item had been shared only linguistically, 3.66 (*SD*= 1.21) when the item had been shared only physically, and 3.47 (*SD*= 1.19) when the item was new. Table 5 shows the means and standard deviations for ratings of the relative clarity of lexically identical expressions according to the information status of the mention (Phase 1 mention and Phase 2 mentions according to the conditions of co-presence), Matcher order and Matcher identity.

Listeners rated expressions that were first mentioned in Phase 1 as clearer than when the expressions were repeated in Phase 2 (Phase 1 vs. all Phase 2 mentions: *F1* (1, 39)= 58.71, $p < .001$; *F2* (1, 76)= 11.76, $p < .01$). But more interesting, when considering expressions mentioned in Phase 2, listeners did distinguish items in their relative clarity ratings depending on how pairs in Experiment 2 had shared them during Phase 1; the co-presence condition mattered, at least by subjects (*F1* (3, 117)= 15.59, $p < .001$; *F2* (1, 82)= .66, *n.s.*[10]). According to listeners' ratings, Directors produced clearer expressions for items that had been shared previously only physically than for those that had been

[10] Since by design an item (triplet) could not have clarity ratings for all four co-presence conditions—the Phase 2 mentions were either a LP-N pair or L-P pair—a main effect of co-presence could not be determined. The effect reported by items here is an interaction between the type of Phase 2 pair (LP-N vs. L-P, a between-items factor) and type of Phase 2 mention (previously mentioned (LP or L), not previously mentioned (P or N)): this interaction indicates whether the co-presence conditions of Phase 1 mattered.

previously shared both linguistically and physically (*F1* (1, 39)= 21.74, $p < .001$; *F2* (1, 76)= 1.45, $p = .23$[11]) or only linguistically (*F1* (1, 39)= 26.26, $p < .001$; *F2* (1, 40)= 1.50, $p = .23$), though these effects were significant only by subjects. They did not reliably produce items that had been shared only linguistically more clearly than those that had been shared linguistically and physically (*F1* (1, 39)= 2.43, $p = .13$; *F2* (1, 76)= .18, *n.s.*). Curiously, mentions of new items were just as attenuated in their clarity as those for items that had been previously shared both linguistically and physically (*F1* (1, 39)= .46, *n.s.*; *F2* (1, 36)= .01, *n.s.*). As seen in Table 5, only for Matcher A in order ABAB did Directors produce referring expressions for new items that were rated as clearer than items that had been shared linguistically and physically (*F1* (1, 39)= 19.29; $p < .001$; *F2* (1, 16)= .06, *n.s.*).

Listeners rated expressions that were produced by Directors in order ABBA as less clear than those in order ABAB (*F1* (1, 39)= 61.09, $p < .001$; *F2* (1, 76)= 2.37, $p = .13$); moreover, expressions to Matcher A in order ABBA were more attenuated compared to Matcher B (*F1* (1, 39)= 13.55, $p < .01$; *F2* (1, 39)= .78, *n.s.*) leading to an interaction of Matcher order and Matcher identity (*F1* (1, 39)= 17.77, $p < .001$; *F2* (1, 76)= 1.51, $p = .22$). This is consistent with a memory-based account for these adjustments, since Directors encountering Matcher A in Round 4 would have had a harder time recalling how they had shared items with them in Round 1, compared to when encountering Matcher B consecutively in Rounds 2 and 3.

[11] For the by-items ANOVAs, the LP vs. P and LP vs. L contrasts have co-presence as a between items factor (since these conditions occur across triplets), whereas the LP vs. N and L vs. P contrasts have co-presence as a within-items factor (since these conditions occur within a triplet). The degrees of freedom in these contrasts reflect this.

The intelligibility results in Experiment 3 provide evidence that Directors in Experiment 2 did adjust their articulation of referring expressions according to *how* the referents had been previously shared. Using relative clarity ratings, I showed partner-specific adjustments, with listeners rating expressions for items that had been shared previously only physically as clearer than those shared only linguistically or both linguistically and physically. These results generalized reliably across listeners, though less so across triplets of excised expressions: this lack of power by items comes with the territory of trying to detect subtle effects in a sample that was selected from a corpus of unconstrained conversations through a "blind," systematic procedure. While increasing the size of the sample in a follow up study would be possible, this could not be done easily without making heavy demands upon listeners: with only 84 triplets, most listeners in this task took between 1.5 to 2 hours to provide ratings. Detecting adjustments in articulation is often difficult in lengthy dialog materials[12]. For example, when articulation is assessed through intelligibility experiments, partner-specific adjustments may not

[12] I conducted an additional experiment in which participants heard only once each of the 84 pairs of referring expressions from Phase 2 (rather than 84 triplets from Phase 1 and 2); after hearing a pair, participants chose which word of the pair was the clearest and indicated their confidence on a 1-4 scale. Listeners were no more likely, or confident, to select an expression for an item that was new for the Matcher in Phase 2 as being clearer than an expression for an item that had been shared linguistically and physically in Phase 1. This is consistent with listeners' ratings for these items in Experiment 3. But unlike listeners' ratings in Experiment 3, which distinguished items that had been shared only physically from those shared only linguistically, participants in this intelligibility experiment were no more likely, or confident, to select an expression for an item that had been shared only physically with a Matcher in Phase 1 as clearer than an expression for an item that had been shared only linguistically. The null results of this experiment can be attributed to the fact that detecting subtle differences in intelligibility was not possible when listeners heard words only once and had to make a forced choice about word clarity. In Experiment 3, enabling listeners to play words as many times as they needed to when comparing them and having listeners give relative ratings as opposed to making an absolute judgment about word clarity, allowed me to uncover the subtle differences between co-presence conditions.

generalize as reliably across referring expressions (Galati & Brennan, 2010), and when it is assessed by measuring word duration, these adjustments may not be detectable at all (Galati & Brennan, 2010; Bard et al, 2000). It is therefore remarkable that I found significant differences in intelligibility that were reliable by subjects for relevant contrasts (namely, for items shared only physically vs. both linguistically and physically, and for items shared only physically vs. only linguistically).

The finding that listeners rated the clarity of new items comparably to items shared linguistically and physically contrasts with some of my earlier work in which expressions repeated to a new partner were rated as more intelligible than those repeated to an old partner and were just as intelligible as first mentions (Galati & Brennan, 2010). However, a critical difference is that in Galati and Brennan (2010) speakers could use their addressees as a global cue for shared information (e.g., *this partner has heard this story before*, or not) in order to make appropriate adjustments in the intelligibility of referring expressions embedded within a story, while in this study speakers could not use their addressees as such a global cue, using them instead as a cue for how referents had been shared on an item-by-tem basis. Because invoking the informational needs of the partner for each referent should be more taxing to working memory than invoking them only once at the beginning of the interaction, Directors in Experiment 2 encountering items that were new for their Matcher may have defaulted to attenuating (or alternatively, not clarifying) their referring expressions if they could not quickly determine how the item had been previously shared. Given the challenge of determining the information status of referents on an item-by-item basis, the constraints of articulation as a fast-acting process may explain why new referents were not distinguished from referents that had

been shared linguistically and physically in their intelligibility; since determining that a referent is new through inference may not be achieved within the permitted time course, Directors may default to attenuating intelligibility of new items in what may be considered to be an egocentric manner. This would suggest that the partner-specific adjustments observed in articulation may be in the direction of clarification as opposed to attenuation: when speakers can determine quickly that they had shared information partially with their partners (only linguistically or only physically), they clarify their pronunciation, but when they cannot determine how they have shared this information quickly enough, they may default to pronouncing words less clearly, much like they do when they determine that they have shared information fully (linguistically and physically) with their partners.

While other studies have suggested that during utterance planning speakers default to clarifying (at least in terms of disambiguating utterances with optional clauses) when they cannot readily cue the informational needs of their partner (Kraljic & Brennan, 2005), it may be the case that for the fast-acting process of articulation speakers default to attenuating. This possibility does not necessitate that partner-specific adjustments engage the language processing system in a modular fashion: although different levels of linguistic representation may have different default parameters (to attenuate vs. to clarify), they still allow for early partner-specific adjustments whenever information about the partner's needs is readily cued.

CHAPTER 9: General Discussion

9.1. Summary of findings

My motives in this dissertation have been twofold. First, I wanted to investigate whether speakers' memory representations for shared information index how information has been previously shared, as reflected by their subsequent partner-specific adjustments. Second, I wanted to extend some of my earlier work (Galati & Brennan, 2010) to determine whether partner-specific adjustments would be observed at the level of articulation as well as at the level of utterance planning—an issue pertinent to whether language processing is modular during audience design.

The findings can be summed up as follows: Directors' partner-specific adjustments, overall, demonstrated sensitivity to how information had been shared, and were related appropriately to the grounding techniques they had previously used with their partners. Directors' descriptions for some behavioral adjustments reflected a binary distinction between information that had been previously mentioned versus not (as reflected in Directors' use of definite expressions), while for other behavioral adjustments the descriptions reflected sensitivity to the specific conditions of co-presence when sharing information (as reflected in their initial words and idea units, and to a lesser extent their hedges and reconceptualizations). When taken together, these adjustments can be thought of as appropriate strategies in initial audience design driven by speakers' memory for how information had been shared with their conversational partners: Referents that had been previously shared both linguistically and physically involved attenuated initial descriptions (fewer words, idea units), fewer markers of provisionality and more markers of definiteness. Referents that had been shared only linguistically were

described with just as many markers of definiteness, signaling to the conversational partner that these referents had been previously mentioned, while at the same time they were described more provisionally and with more detail, with just as many words and idea units as completely new referents, reflecting the degree of grounding that the conversational partners had previously achieved. For these referents, conversational partners had previously exchanged evidence in grounding until Matchers ascertained with enough certainty for current purposes that they did not have them, but the exact identity of these referents remained unknown until Phase 2. Indeed, consistent with other studies showing that the process of grounding is shaped by whether speakers have visual evidence of their addressees' understanding (Brennan, 1990, 2005; Clark & Krych, 2004; Gergle, Kraut & Fussell, 2004), when discussing referents that the Matchers didn't have in Phase 1 Directors' cumulative descriptions involved more words, idea units, hedges and metacomments, and also conversational partners' exchanges extended over more turns, involving more errors and more requests by Matchers for a clarification or expansion of the description. Thus, the Directors' reintroduction of these referents in Phase 2 with enough detail, yet with definite expressions marking them as already mentioned suggests that Directors did in fact remember how these referents had been shared.

These findings complement well the findings of a recent set of studies by Brown-Schmidt (2009b) who found that people can actually keep track of whether grounding was completed or aborted during referential communication. That is, upon hearing a temporarily ambiguous question (e.g., "What's above the cow that's wearing shoes?" in a context with two cows), addressees were more likely to direct attention away from an

object (above one of the cows) that had been previously mentioned and grounded and toward the unmentioned target (above the other cow); critically, this effect was attenuated when the target object had been mentioned but not grounded. Both my and Brown-Schmidt's (2009b) studies highlight that people keep track of not only whether they've discussed something before, but also whether they've grounded it or not.

Moreover, Directors here distinguished referents that had been previously shared only linguistically vs. only physically: initial descriptions for referents that had been shared only physically included more words and idea units, more reconceptualizations, more hedging and fewer definite expressions. Referents that had been shared only physically in Phase 1 were treated in some ways just as completely new referents (e.g. with comparable amounts of definite expressions and reconceptualizations), though in other ways they were treated as even *newer* then new referents (e.g., involving more words, idea units, and even referring expressions that were rated as clearer by listeners)! Evidence from the Directors' source monitoring questionnaires suggests that they indeed showed more sensitivity at explicitly remembering referents that had been previously shared only physically as opposed to being absent from Phase 1. This is consistent with the idea that determining that a referent had been previously shared physically involves accessing an existing memory trace for that experience, while determining that an item had not been shared at all is done by inference, by failing to access an episodic trace for that partner-item association. Accessing a memory trace for a referent that had not been described before (though it had been shared physically) perhaps constitutes more compelling evidence for Directors to elaborate their descriptions over inferring that a referent had not been shared before. In this light, the fact that Directors were faster to

begin their descriptions for referents that had been shared linguistically and physically than for those that had been shared partially or not at all may reflect the processing cost of selecting an appropriate grounding strategy instead of a failure to distinguish between these items.

Remarkably, speakers make distinctions in their articulation according to the conditions of co-presence, as demonstrated in Experiment 3. A comparison of Experiments 1 and 2 here is pertinent: In Experiment 1, target items were harder to lexicalize than in Experiment 2, and as a result conversational partners spent more effort negotiating descriptions for tangrams than for category exemplars. But the fact that Directors' partner-specific adjustments in articulation were also sensitive to the conditions of co-presence of Phase 1, in conjunction with Directors' distinctions in their explicit recall, suggests that the findings of Experiment 1 cannot be attributed only to the elaborated grounding process that tangrams involved. Regardless of any effect grounding might have had on episodic traces for shared information, the conditions of co-presence in both Experiments 1 and 2 were represented robustly enough in Directors' episodic traces to enable partner-specific adjustments at the levels of both utterance planning and articulation. These findings demonstrate that during audience design the language processing system is engaged in a non-modular fashion: in line with some of my earlier work (Galati & Brennan, 2010), the fast-acting process of articulation is not encapsulated from partner-specific knowledge.

9.2. Implications for models of audience design, language processing, and memory

These results have implications for several areas of research, including audience design, language processing models, and memory. First of all, they demonstrate that

people can keep track of simple distinctions about their conversational partners' informational needs within a reasonably complex communicative setting, even if these distinctions are not necessarily binary. Although in many experimental studies demonstrating audience design effects the conversational partners' informational needs could be represented with a single, binary distinction or constraint, my dissertation shows evidence that speakers can keep track of more than a single constraint pertinent to a shared experience (or alternatively, they can keep track of a multiple-valued constraint). As reflected by the gradient adjustments on various dimensions of speech planning (e.g., words, idea units, intelligibility of referring expressions), explicit recall, and the strategies speakers adopted overall (e.g., choosing to describe referents that had been shared only linguistically less provisionally but with more detail), speakers here distinguished whether information had been shared linguistically and physically, only linguistically, or only physically. Speakers could achieve this either by computing two constraints–one about whether information had been previously mentioned vs. not, and one about whether it had been previously shared physically vs. not—or by computing a multiple valued constraint distinguishing the different co-presence conditions. In any case, speakers can evidently keep track of how information had been shared as long as the distinction is simple enough and can be cued rapidly.

Keeping track of the conditions of co-presence or the affordances of a prior shared communicative situation is beneficial for conversational partners because it enables them to gauge the appropriate amount of time and effort they need to expend in grounding in subsequent conversations. Thus, not only do the current affordances of a communicative situation affect how people ground their conversations through ongoing

evidence about what they do or do not understand (Clark & Brennan, 1991; Brennan, 1990, 2005; Clark & Krych, 2004; Gergle, Kraut & Fussell, 2004), but the affordances at the encoding of shared information can affect how speakers later design their initial descriptions and how they coordinate with their partners. My findings here run counter to the idea proposed by Pickering and Garrod (2004) that conversational partners achieve converging mental representations without tracking anything specific about the information they've shared with them, using instead simple cognitive mechanisms (such as priming) during initial processing and adapting to their partners by using their own knowledge as a proxy. As shown, people in fact are capable of tracking what information they have shared with whom, and moreover under what conditions they have shared it. Moreover, the speakers' adjustments here cannot be accounted for in terms of an automatic, "dumb" adjustment like priming; they are best interpreted in terms of selecting appropriate grounding strategies depending on how information had been previously shared.

Another implication of this work is that inferences about the informational needs of conversational partners, drawn from simple and rapidly cued memory representations for the conditions of co-presence, result in audience design effects at multiple grains of spontaneous language processing. These include a grain assumed to be too automatic to show audience design effects: articulation. I did not find evidence of any "dual process" during audience design, whereby articulation is encapsulated from partner-specific knowledge and defaults to egocentric processing, while more resource-consuming processes like utterance planning are guided by inferences about the partner's needs (Bard et al., 2000; Bard & Aylett, 2001). Speakers in Experiment 2 faced a remarkable

challenge, since they had to determine the informational needs of their partners on a referent-by-referent basis and could not rely on a single global cue to reconstruct those needs. Compounded with this challenge, the constraints of articulation as a fast-acting process explain perhaps why new referents were not distinguished from referents that had been shared linguistically and physically in their intelligibility here; since determining that a referent is new through inference may not be achieved within the permitted time course, Directors may default to attenuating intelligibility of new items in what may be considered to be an egocentric manner. But, critically, determining the information status of partially shared referents (linguistically-only and physically-only) can be achieved within the time constraints of the articulation process, as these referents are distinguished from fully shared items and from one another. In other words, as long as the information status of a referent can be easily cued, even fast-acting processes like articulation are nimble enough to be impacted by prior shared experiences with a conversational partner. Therefore, initial processing need not be egocentric (as proposed by Brown & Dell, 1987; Keysar, Barr, Balin, & Paek, 1998; Kronmuller & Barr, 2007) for any stage of language planning. The speech planning system, which is characterized by incremental processing and a cascading architecture (e.g., Levelt, 1989; Dell, 1986; Dell & O'Seaghdha, 1992; Bock, 1995; Bock & Levelt, 1994), is flexible and can implement partner-specific adjustments in a non-modular fashion at various levels of planning.

Altogether, my findings lend insight both into how shared information can be indexed in episodic memory traces and into how speakers, upon accessing these memory traces using their conversational partners as cues, can rapidly perform adjustments

appropriate to their partners' informational needs across levels of the language processing architecture.

9.3. Future research directions

In this dissertation I have demonstrated that speakers can keep track of at least two simple constraints regarding common ground (or a single constraint with multiple values). I am interested in exploring the limits of representing and keeping track of such constraints. In a follow up study, I would be interested in extending the current design by manipulating further dimensions regarding the conversational partner's informational needs in order to see whether speakers can successfully compute these constraints during early processing and perform appropriate partner-specific adjustments. For example, in addition to manipulating the conditions of co-presence under which items have been shared, I could vary characteristics of the addressees, such as whether they are native speakers of English or not. Insofar as additional constraints are relevant to the task at hand, I would like to see whether and when they become readily available and how they may interact with other constraints about common ground.

One surprising finding in this dissertation was that speakers elaborated their descriptions of items shared only physically more so than completely new items. This finding may have arisen from the fact that the conditions of co-presence were complementary across Matchers in Phase 1: items shared only physically may have been more distinctive than new items just by virtue of being shared only linguistically with the other partner. Therefore, it would be worthwhile to replicate the present study while avoiding this complementary distribution of co-presence conditions, by uniquely associating items with only one co-presence condition for a single conversational partner.

Such a design would allow me to determine whether items that have been shared only physically are, in general, more distinctive than new items—perhaps because they involve actual episodic traces as opposed to inference—or whether their distinctiveness was specific to this dissertation's design.

Another research direction relevant to understanding the limits of assessing constraints about common ground, involves understanding better the role of executive control. Executive control has been shown to be critical to audience design; failures in inhibiting one's own perspective can account for some occasional failures to consider shared information (Brown-Schmidt, 2009b). For people to select the appropriate perspective (or suppress the inappropriate one) during audience design, executive control is required (Vogeley et al., 2001; Brass, Derrfuss, Forstmann, & von Cramon, 2005). Manipulations of executive control often involve having participants engage in a secondary task, but another approach is to manipulate how automatic the association is between a conversational partner and shared information, potentially with different types of category cues. This idea follows Horton and Gerrig's (2005a) suggestion that processing is more strategic (and presumably involves more executive control) when common ground is less obvious and associated with less-evident category cues (e.g., establishing whether a conversational partner is a native New Yorker might involve a more strategic assessment of community membership than establishing whether they are a native English speaker). On the other hand, processing may be more automatic when common ground is indexed by clear category cues; there is evidence, for example, that cues for community membership categories are processed automatically and unconsciously, influencing social perception, social judgments, and stereotype use

(Bargh, 1996). Processing common ground that is associated with an *ad hoc* category (i.e., a category grouping created spontaneously for use in specialized contexts) should be then more strategic and less automatic than having a common category cue. This idea is consistent with findings that ad hoc categories are not pre-stored in memory as opposed to common categories (e.g., there is less consistency in producing exemplars for ad hoc categories than common categories), even though ad-hoc categories have properties (e.g. typicality gradients in ratings) that are as salient as those of common categories (Barsalou, 1983). These distinct properties for common and ad hoc categories could lend themselves well to exploring whether people assess common ground with different degrees of recruiting executive control, depending on the kind of category for which the conversational partner serves as a cue. Behavioral studies can investigate whether linguistic adjustments reflect differences in the ease of assessing shared information depending on whether partners serve as a cue for a common category, an ad hoc category, or no category.

My findings in this dissertation are in line with the view that audience design need not follow from accessing specialized representations for shared information, but rather from accessing episodic traces through ordinary memory processes (Horton & Gerrig, 2002, 2005a, 2005b). Yet a complete account of audience design would also involve understanding the neural underpinnings of partner-adapted processing; specifically, how this processing recruits neural circuits that may handle a variety of pertinent functions, including language processing, monitoring nonverbal cues, memory processes, and "theory of mind" or mentalizing (see Brennan, Galati, & Kuhlen, to appear, for a review of relevant findings). The ability to have a theory of mind is critical to audience design as

it involves being able to understand the behavior of self and others by attributing independent mental states. Areas activated during memory processes have been implicated in neural circuits supporting mentalizing and the processing of social cues, like eye gaze, bodily movement and orientation (Gallagher & Frith, 2003; Overwalle & Baetens, 2009). Since audience design involves accessing episodic traces for shared information or semantic memories to generate scripts and stereotypes (relevant to assessing community membership), it would be interesting to investigate whether audience design activates the same areas found active during episodic memory retrieval (e.g., when recollecting familiar faces and scenes, Nakamura et al, 2000; recognizing familiar voices, Nakamura et al, 2001; and retrieving autobiographical memories, Fink et al., 1996) or semantic memory retrieval (Funnel, 2001). In addition to investigating the involvement of memory-related areas, future research could examine how neural circuits supporting audience design and mentalizing interact, as proposed by Brennan, Galati, & Kuhlen (to appear). To the extent that functional imaging studies provide evidence that during audience design areas involved in memory processes are activated along with other areas involved in mentalizing, perhaps a qualifier should be appended to the memory-based view of audience design (Horton & Gerrig, 2002, 2005a, 2005b; Metzing & Brennan, 2003): although audience design may be achieved by accessing general, not specialized, episodic or semantic memories, it unfolds within a neural circuit that *is* specialized.

This dissertation contributes to our understanding of what aspects of common ground speakers encode in memory representations for shared information, and how they recruit their language processing system to deploy appropriate partner-specific

adjustments at multiple grains of linguistic representation. The results highlight just how flexibly speakers can adapt their behavior, incorporating whatever information is available about their shared experience with a partner into their early utterance planning. But processes involved in audience design are yet to be fully understood, as they involve different subprocesses interacting in a highly integrated manner, including language processing, memory retrieval, mentalizing, executive control, and monitoring the partner's attention and action. The enterprise of understanding audience design and, more broadly, the coordination of the behavior and mental representations of conversational partners should pursue further how these processes interact (e.g., the role of executive control in mentalizing; how speakers update the model of their addressees' needs based on incoming feedback). Along with behavioral evidence, timing and anatomical evidence from eye-tracking, electrophysiological and imaging studies will be instrumental in unveiling how these processes interact.

References

Anderson, J. R. & Bower, G. H. (1973). *Human associative memory.* Washington, DC: Winston.

Arnold, J. E., Tanenhaus, M. K , Altmann, R. & Fagnano, M. (2004). The old and thee, uh, new. *Psychological Science, 15*, 578-581.

Balota, D.A., Yap, M.J., Cortese, M.J., Hutchison, K.A., Kessler, B., Loftis, B., Neely, J.H., Nelson, D.L., Simpson, G.B., & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, *39*, 445-459.

Bargh, J. A. (1996). Automaticity in social psychology. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 169-183). New York: Guilford.

Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., & Newlands, A. (2000). Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory & Language*, *42*, 1-22.

Bard, E. G., & Aylett, M. P. (2001). Referential form, word duration, and modeling the listener in spoken dialogue. Paper presented at the Twenty-third Annual Conference of the Cognitive Science Society, Edinburgh, Scotland.

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

Bavelas, J.B., Chovil, N., Coates, L., and Roe, L. (1995). Gestures specialized for dialogue. *Personality and Social Psychology Bulletin*, *21*, 394-405.

Bock, J. K. (1995). Speech production: From mind to mouth. In J. L. Miller & P. D. Eimas (Eds.), *Speech, language, and communication* (pp. 181-216), San Diego, CA: Academic Press

Bock, J. K., & Levelt, W.J.M. (1994). Language production: Grammatical encoding. In M. A. Gernsbacher (Ed.) *Handbook of psycholinguistics* (pp. 945-984). San Diego, CA: Academic Press

Bortfeld, H., & Brennan, S. E. (1997). Use and acquisition of idiomatic expressions in referring by native and non-native speakers. *Discourse Processes*, 23, 119-147.

Brass, M., Derrfuss, J., Forstmann, B. & von Cramon, D. Y. (2005). The role of the inferior frontal junction area in cognitive control. *Trends in Cognitive Science*, *9* (*7*), 314-316.

Brennan, S. E. (1990). Seeking and Providing Evidence for Mutual Understanding. Unpublished doctoral dissertation, Stanford University.

Brennan, S. E. (2005). How conversation is shaped by visual and spoken evidence. In J. Trueswell & M. Tanenhaus (Eds.), *Approaches to studying world-situated language use: Bridging the language-as-product and language-action traditions* (pp. 95-129). Cambridge, MA: MIT Press.

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.

Brennan, S. E., Galati, A., & Kuhlen, A. K. (to appear). Two minds, one dialog: Coordinating speaking and understanding. *Psychology of Learning and Motivation*, *53*.

Brennan, S. E., & Hanna, J. E. (2008). Partner-specific adaptation in dialogue. *Journal of Memory and Language.*

Brennan, S. E., & Schober, M. F. (2001). How listeners compensate for disfluencies in spontaneous speech. *Journal of Memory and Language*, *44*, 274-296.

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

Bock, J. K. (1986). Meaning, sound, and syntax: Lexical priming in sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*, 575-586.

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

Brown-Schmidt, S. (2009a). Partner-specific interpretation of maintained referential precedents during interactive dialog. *Journal of Memory and Language, 61*, 171-190.

Brown-Schmidt, S. (2009b). The role of executive function in perspective taking during online language comprehension. *Psychonomic Bulletin and Review*, *16* (*5*), 893-900.

Brugman, H., & Russel, A. (2004). Annotating Multi-media / Multi-modal resources with ELAN. *LREC.*

Cahn, J.E., & Brennan, S. E. (1999). A psychological model of grounding and repair in dialog. *Proceedings, AAAI Fall Symposium on Psychological Models of Communication in Collaborative Systems.* (pp. 25-33). North Falmouth, MA: American Association for Artificial Intelligence.

Carney, R. N., & Levin, J. R. (2002). Pictorial illustrations still improve students' learning from text. *Educational Psychology Review*, *13* (*1*), 5-26.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. Resnick, J. Levine, & S. Teasley, (Eds). *Perspectives on socially shared cognition* (pp. 127-149). Washington, DC: APA Books.

Clark, H. H., & Carlson, T. B. (1982). Hearers and speech acts. *Language*, *58*, 332-373

Clark, H. H., & Havilland, S.E. (1977). Comprehension and the given-new contract. In: R. O Freedle (Ed.), *Discourse production and comprehension*, (pp. 1-40). Norwood, NJ: Ablex Publishing Corporation.

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

Clark, H. H., & Marshall, C. E. (1981). Definite reference and mutual knowledge. In A. J. Jodhi, B. L. Webber, & I. A. Sag (Eds.), *Elements of discourse understanding*, (pp. 10-63). Cambridge: 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*. Amsterdam: North-Holland.

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

Cohen, A. & Harrison, R.P. (1973). Intentionality in the use of hand illustrators in face-to-face communication situations. *Journal of Personality and Social Psychology*, *28*, 276-279.

Cooper, W. E., & Paccia-Cooper, J. (1980). Syntax and speech. Cambridge, MA: Harvard University Press.

Dell. G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93* (*3*), 283-321.

Dell. G. S., & O'Seaghdha, P. G. (1992). Stages of lexical access in language production, *Cognition*, *42*, 287-314.

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

Fink, G. R., et al. (1996). Cerebral representation of one's own past: neural networks involved in autobiographical memory. *Journal of Neuroscience*, *16*, 4275-4282.

Funnel, E. (2001). Evidence for scripts in semantic dementia: Implications for theories of semantic memory. *Cognitive Neuropsychology*, *18*, 323-341.

Galati, A., & Brennan, S. E. (2010). Attenuating information in Spoken Communication: For the speaker, or for the addressee? *Journal of Memory and Language*, *62*, 35-51.

Gallagher, H. L., & Frith, C. D. (2003). Functional imaging of "theory of mind". *Trends in Cognitive Sciences*, *7* (*2*), 77-83.

Gergle, D., Kraut, R. E., Fussell, S. R. (2004). Language efficiency and visual technology: Minimizing collaborative effort with visual information. Journal of Language and Social Psychology, *23* (*4*), 491-517.

Gibson, J. J. (1996). *The senses considered as perceptual systems.* Boston, MA: Houghton Mifflin.

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.

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

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

Horton, W. S., (2007). The influence of partner-specific memory associations on language production: Evidence from picture naming. *Language and Cognitive Processes*, *22* (*7*), 1114-1139.

Horton, W. S., & Gerrig, R. J. (2005a). Conversational common ground and memory processes in language production. *Discourse Processes*, *40*, 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., & Gerrig, R. J. (2002). Speaker's experiences and audience design: knowing when and knowing how to adjust utterances to addressees. *Journal of Memory and Language*, *47*, 589-606.

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

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

Johnson, M.K., Hashtroudi, S., & Lindsay, D.S. (1993). Source monitoring. *Psychological Bulletin, 114,* 3-28

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

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

Keysar, B., Barr, D. J., & Horton, W. S. (1998). The egocentric bias of language use: Insights from a processing approach. *Current Directions in Psychological Science*, *7*, 46-50.

Krauss, R. M., & Weinheimer, S. (1964). Changes in reference phrases as a function of frequency of usage in social interaction: A preliminary study. *Psychonomic Science*, *1*, 113-114.

Kronmüller, E., & Barr, D. J., (2007). Perspective-free pragmatics: Broken precedents and the recovery-from-preemption hypothesis. *Journal of Memory and Language, 56*, 436–455.

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

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

Lockridge, C. B (2007). Impact of working memory span on referring in conversation. *Unpublished doctoral dissertation*. State University of New York: Stony Brook, NY.

MacDonald, M. C. (1994). Probabilistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes, 9*, 157–201.

MacMillan, N. A., & Creelman, C. D. (1991). *Detection Theory: A User's Guide.* New York: Cambridge University Press.

McRae, K, Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language*, *38*, 283 -312.

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

Murray, M. M., Michel, C.M., Grave de Peralta Menendez, R., Ortigue, S., Brunet, D., Gonzalez Andino, S. L., & Schnider, A. (2004). Rapid discrimination of visual and multisensory memories revealed by electrical neuroimaging, *Neuroimage, 21*, 125-135.

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

Nakamura, K. et al. (2000). Functional delineation of the human occipito-temporal areas related to face and scene processing: a PET study. *Brain*, *123*, 1903-1912.

Nakamura, K. et al. (2001). Neural substrates of familiar voices: a PET study. *Neuropsychologia*, *39*, 1047-1054.

Paivio, A. (1986). *Mental representations: A dual coding approach.* Oxford, England: Oxford University Press.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*, 167-226.

Polichak, J. W., & Gerrig, R. J. (1998). Common ground and everyday language use: comments on Horton and Keysar (1996). *Cognition*, *66*, 183-189.

Russell, A. W., & Schober, M. F. (1999). How beliefs about a partner's goals affect referring in goal-discrepant conversations. *Discourse Processes*, *27*(*1*), 1-33.

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

Snodgrass, J. G., & Vanderward, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6* (*2*). 174-215.

Stalnaker, R. (1974). Pragmatic Presuppositions. In M. K. Munitz & P. K. Unger (eds.), Semantics and Philosophy (pp. 197-213). New York University Press, New York. University Press

Swets, B., Jacovina, M., and Gerrig, R. J. (in revision). Sentence production in monologue and dialogue: The scope of planning.

Szekely, A., D'Amico, S., Devescovi, A., Federmeier, K., Herron, D., Iyer, G., Jacobsen, T., & Bates, E. (2003) Timed picture naming: Extended norms and validation against previous studies, *Behavior Research Methods, Instruments & Computers*, *35* (*4*), 621-633.

Tanenhaus, M. K., & Trueswell, J. C. (1995). Sentence comprehension. In J. Miller, & P. Eimas (Eds.), *Handbook of cognition and perception*. San Diego, CA: Academic Press.

The International Picture Naming Project at CRL-UCSD (n.d.) Retrieved November 15, 2009, from http://crl.ucsd.edu/~aszekely/ipnp/

Van Overwalle, F. & Baetens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: A meta-analysis, *Neuroimage*, *48*, 564-584.

Wilkes-Gibbs, D. (1986). *Individual goals and collaborative actions*. Unpublished doctoral dissertation, Stanford University, Stanford, CA.

Appendix A
List of stimuli (tangrams) of Experiment 1

Set 1

item 1.1.

item 1.2

item 1.3

item 1.4

item 1.5

item 1.6

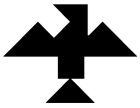
item 1.7

item 1.8

Set 2

item 2.1

item 2.2.



item 2.3

item 2.4

item 2.5

item 2.6

item 2.7

item 2.8

Appendix B

Example of coding idea units. Director 30 describes item 1.8 to Matcher A in Phase 1.

According to the transcription conventions, pauses are marked /. Fillers are annotated as <uh> or <um>. The end point of a self-interruption is marked with an asterisk (*), and curly brackets {} contain undecipherable speech**.** Metacomments are in italics, and are not taken to contribute to propositional content. Instrumental actions are in square brackets [ ].

In this example the idea units of the initial description as the same as those for the total description, since after her first turn the Director only makes a metacomment.

Round 1 (Phase 1) to Matcher A:

D30: okay
this looks like /
a<a> / ballerina /
on a triangle*
upside down triangle / stage /
<um> it's doing a pirouette /
or I don't know what those things are /
%laugh

M_A: %laugh
I don't think I have it /
[*puts up joker card*]

D30: *even if I knew {I…}*
%laugh

M_A: %laugh

Idea units of initial description (7): 1. ballerina 2. on triangle 3. upside down (triangle stage) 4. (upside down) triangle (stage) 5. (upside down triangle) stage 6. doing (pirouette) 7. (doing) pirouette

Idea units of total description (7): 1. ballerina 2. on triangle 3. upside down (triangle stage) 4. (upside down) triangle (stage) 5. (upside down triangle) stage 6. doing (pirouette) 7. (doing) pirouette

Appendix C

Examples of coding for reconceptualizations. The first example illustrates how reconceptualizations were coded when Director 30 described item 1.8 to Matcher B in Round 3 (Phase 2). The second example illustrates how reconceptualizations were coded when she described the same item to Matcher A in Round 4 (Phase 1), to whom she had described this item earlier in Phase 1 (see Appendix B).

To code for reconceptualizations, the idea units of the initial description in a given Phase 2 round are first determined, and then compared to the total idea units of a description of the same item in Phase 2. The comparison involves identifying which idea units in Phase 2 overlap with Phase 1, which are unique to Phase 2, and which are unique to Phase 1. Reconceptualizations are considered to be the idea units that are unique to Phase 2.

Round 3 (Phase 2) to Matcher B:

D30: okay
this look like
somebody / <um>
twirling on a st* <uh>
triangle s*
upside | down triangle stage ? /
with one leg up in the air | /
{on} pointy toes /

M_B: okay /
this<s> ? /
[*puts up standing man card*]

D30: yep

Idea units of initial description (11): 1. somebody 2. twirling 3. upside down (triangle stage) 4. (upside down) triangle (stage) 5. (upside down triangle) stage 6. one (leg) 7. (one) leg 8. (one leg) up (in the air) 9. (one leg up) in the air 10. pointy (toes) 11. (pointy) toes

Idea units overlapping between Phase 1 and Phase 2 (3): 1. upside down (triangle stage) = upside down (triangle stage) 2. (upside down) triangle (stage) = (upside down) triangle (stage) 3. (upside down triangle) stage = (upside down triangle) stage

Idea units unique to the initial description of Phase 2 (Reconceptualizations) (8): 1. somebody 2. twirling 3. one (leg) 4. (one) leg 5. (one leg) up (in the air) 6. (one leg up) in the air 7. pointy (toes) 8. (pointy) toes

Idea units unique to Phase 1 (4): 1. ballerina 2. on triangle 3. doing (pirouette) 4. (doing) pirouette

<u>Round 4 (Phase 2) to Matcher A:</u>

D30: this is the ballerina /
%laugh
M_A: %laugh
[*puts up standing man card*]
D30: on the stage /

Idea units of initial description (1): 1. ballerina

Idea units overlapping between Phase 1 and Phase 2 (1): 1. ballerina

Idea units unique to the initial description of Phase 2 (Reconceptualizations) (0): 0

Idea units unique to Phase 1 (6): 1. on triangle 2. upside down (triangle stage) 3. (upside down) triangle (stage) 4. (upside down triangle) stage 5. doing (pirouette) 6. (doing) pirouette

Appendix D

The 16 target items (category exemplars) of Experiment 2, along with their associated category, visual complexity score (Szekely et al., 2003) and log HAL frequency (from Balota et al, 2007).

Item	Category	Visual Complexity	log Hal frequency
Set 1			
1.1. spider	animal	37059	8.847
1.2. motorcycle	vehicle	24207	8.215
1.3. carrot	food	13201	7.269
1.4. sweater	clothing	11622	7.181
1.5. guitar	instrument	12032	10.118
1.6. closet	furniture	30610	8.568
1.7. finger	body part	5370	10.534
1.8. flashlight	tool	15410	6.739
Set 2			
2.1. butterfly	animal	24645	7.494
2.2. helicopter	vehicle	18241	8.071
2.3. pineapple	food	20721	6.457
2.4. stocking	clothing	16152	7.268
2.5 trumpet	instrument	13615	8.876
2.6. mirror	furniture	11938	10.208
2.7. shoulder	body part	6274	9.345
2.8 scissors	tool	13042	7.222

Appendix E
Figures 1 through 8

Figure 1
Arrangement of Directors and Matchers in a Phase 1 round of Experiments 1 and 2. Phase 2 rounds involved the same arrangement, except the target area was a board with eight slots.

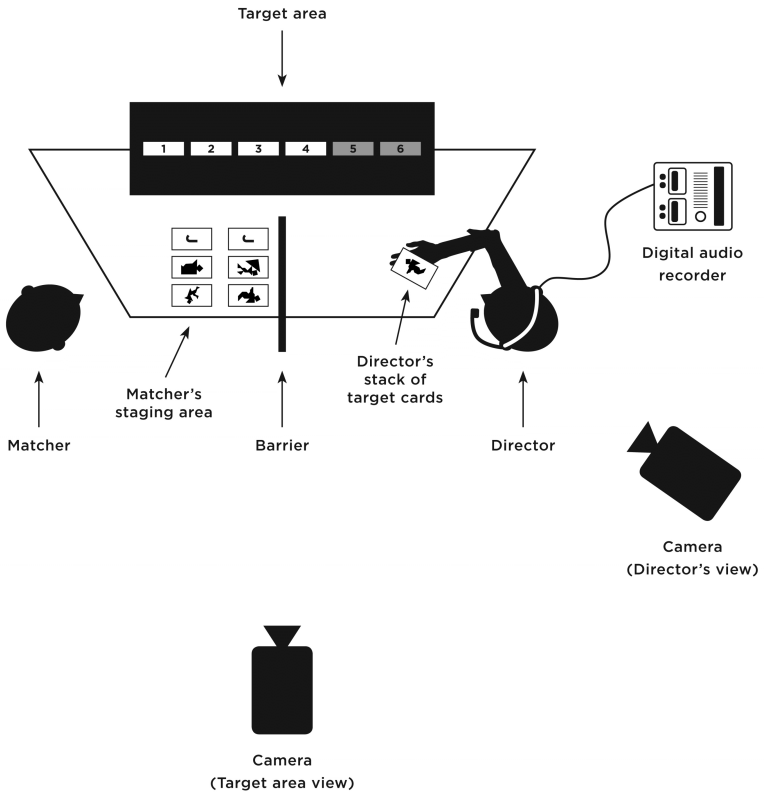


Figure 2
Experiment 1, Mean Proportions of Tangram Items with a Definite Expression or Meta-Reference in the Directors' Initial Description in Phase 2, according to the Co-Presence Conditions of Phase 1 and Matcher Identity. Bars represent standard errors.

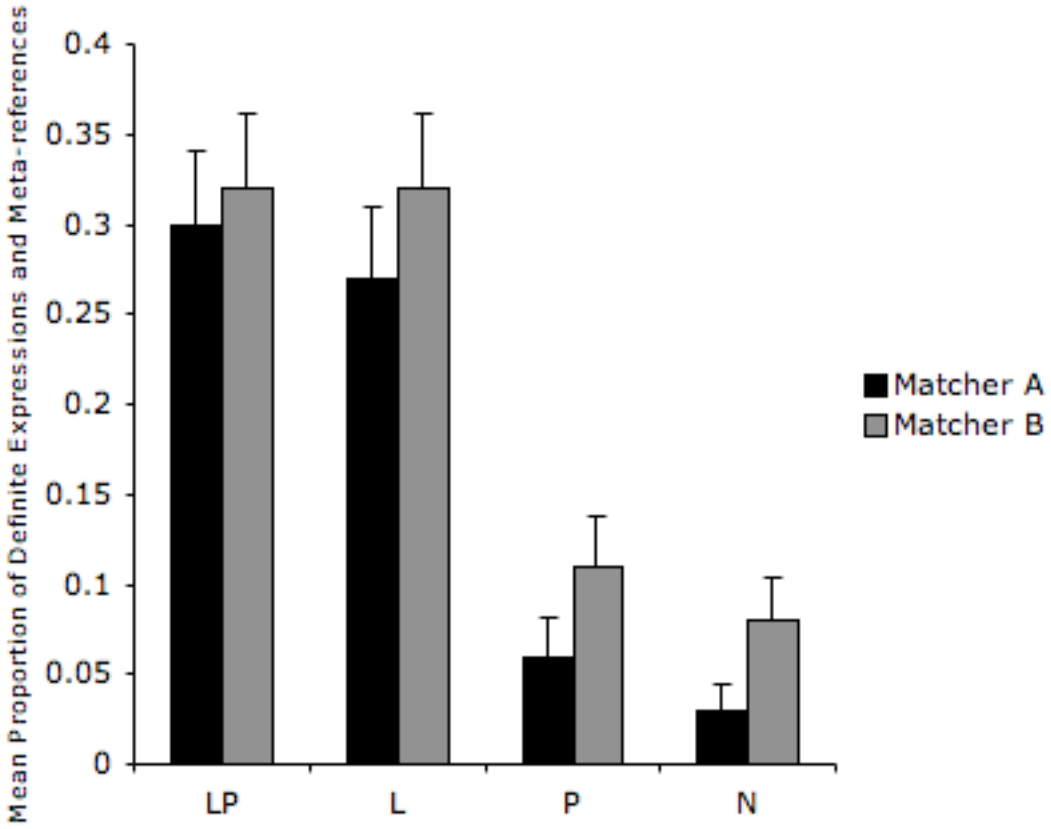


Figure 3
Experiment 1, Mean Number of words in the Directors' Initial Description in Phase 2, according to the Co-Presence Conditions of Phase 1 and Matcher Identity. Bars represent standard errors.

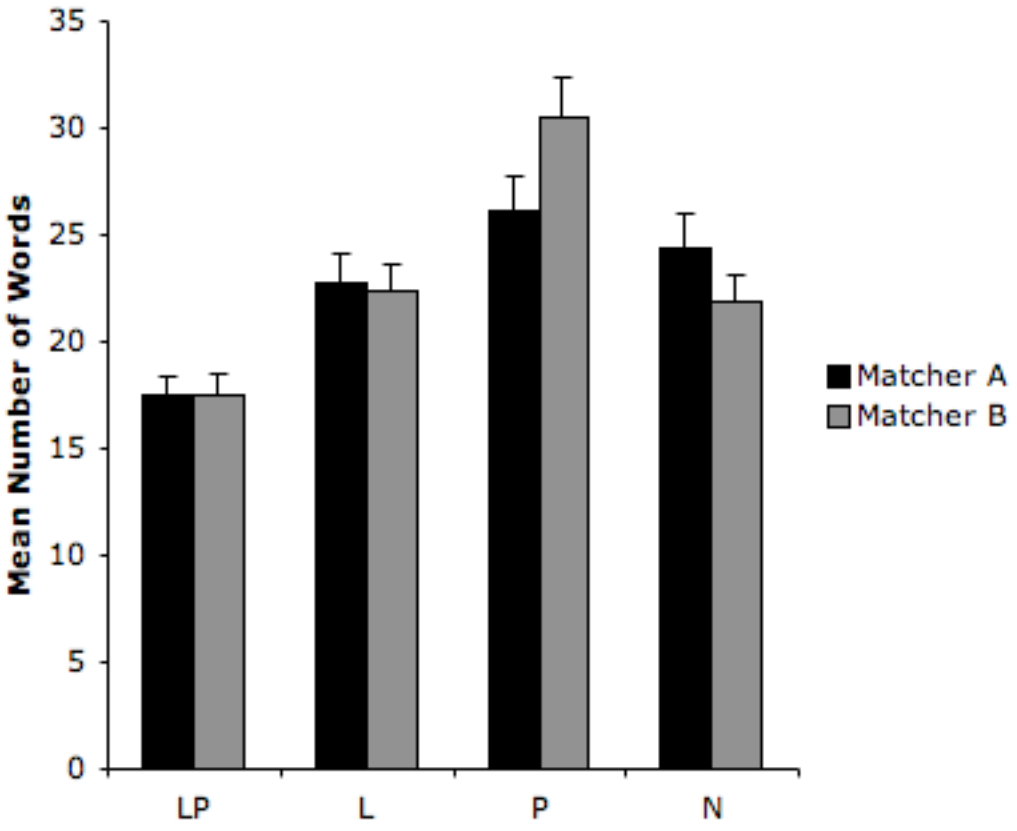


Figure 4
Experiment 1, Mean Number of Idea Units in the Directors' Initial Description in Phase 2, according to the Co-Presence Conditions of Phase 1 and Matcher Identity. Bars represent standard errors.

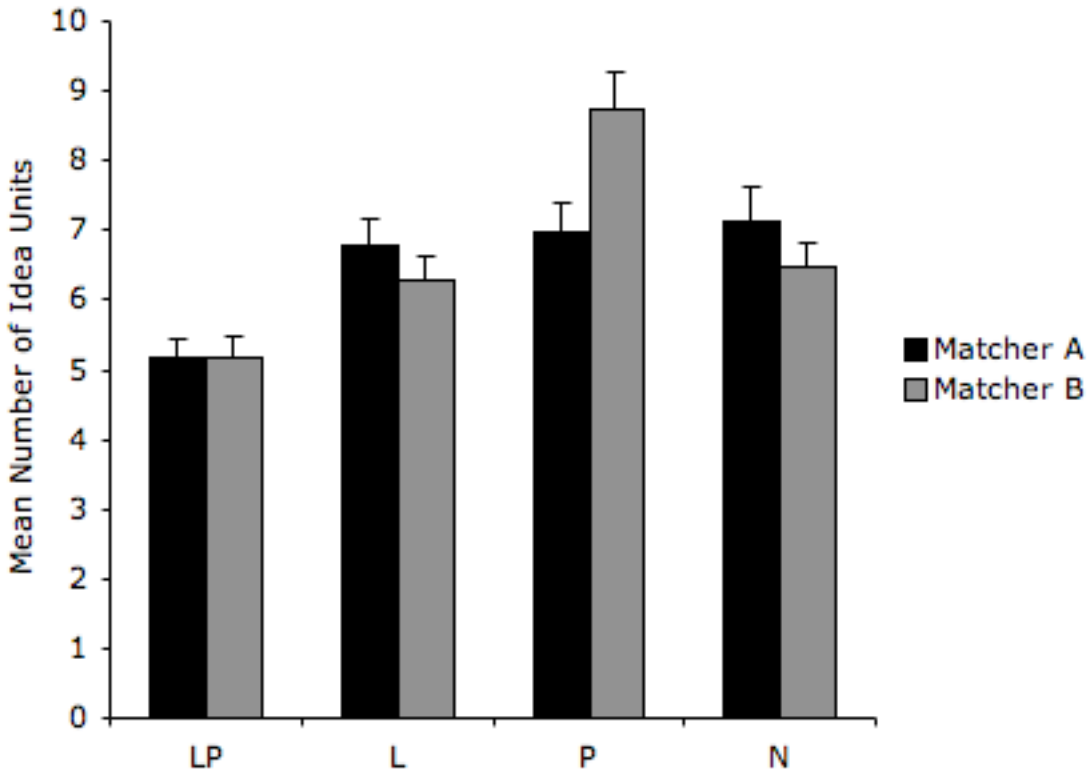


Figure 5
Experiment 1, Mean Number of Reconceptualizations in the Directors' Initial Description in Phase 2, according to the Co-Presence Conditions of Phase 1 and Matcher Identity. Bars represent standard errors.

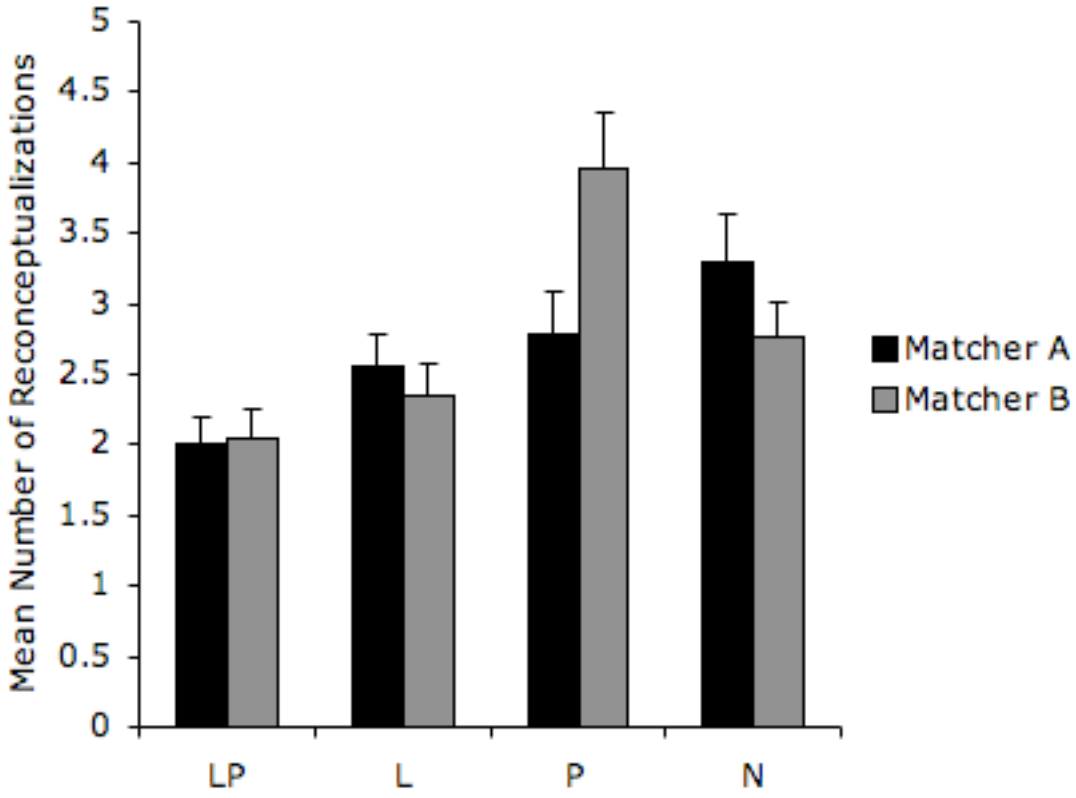


Figure 6
Experiment 1, Mean Proportions of Tangram Items with Hedges in the Directors' Initial Description in Phase 2, according to the Co-Presence Conditions of Phase 1 and Matcher Identity. Bars represent standard errors.

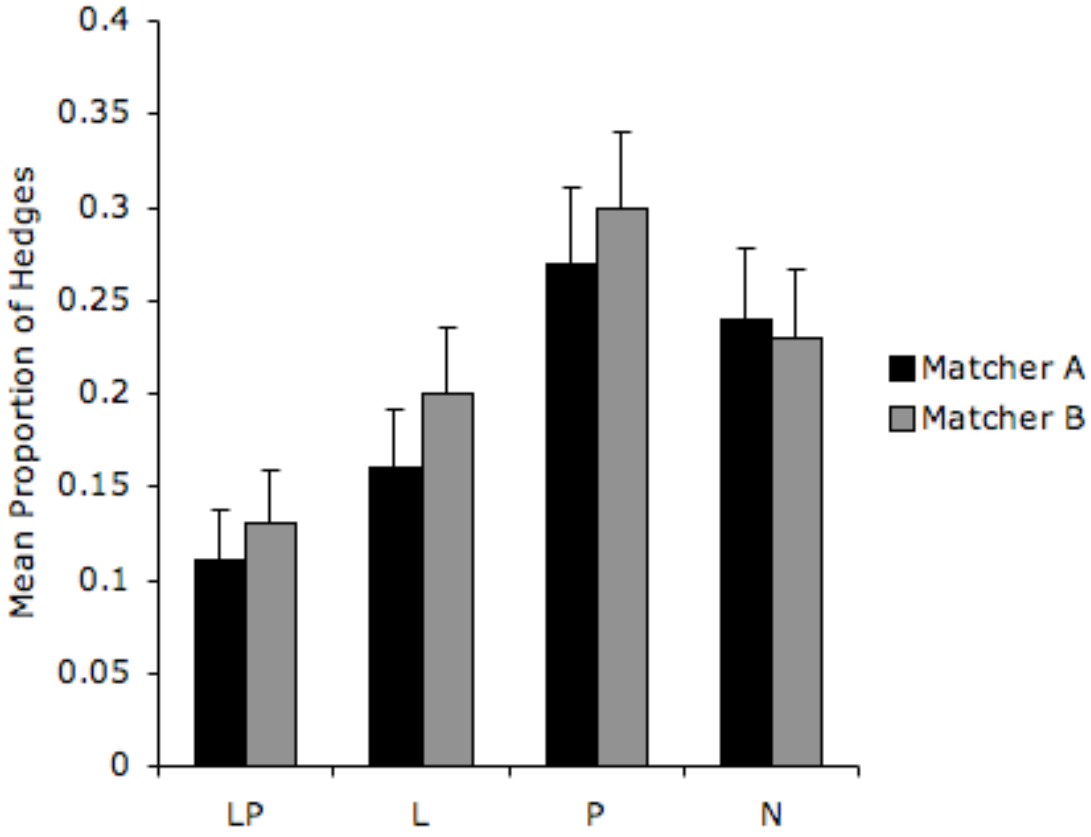


Figure 7
Experiment 1, Mean Proportions of Tangram Items with Metacomments in the Directors' Initial Description in Phase 2, according to the Co-Presence Conditions of Phase 1 and Matcher Identity. Bars represent standard errors.

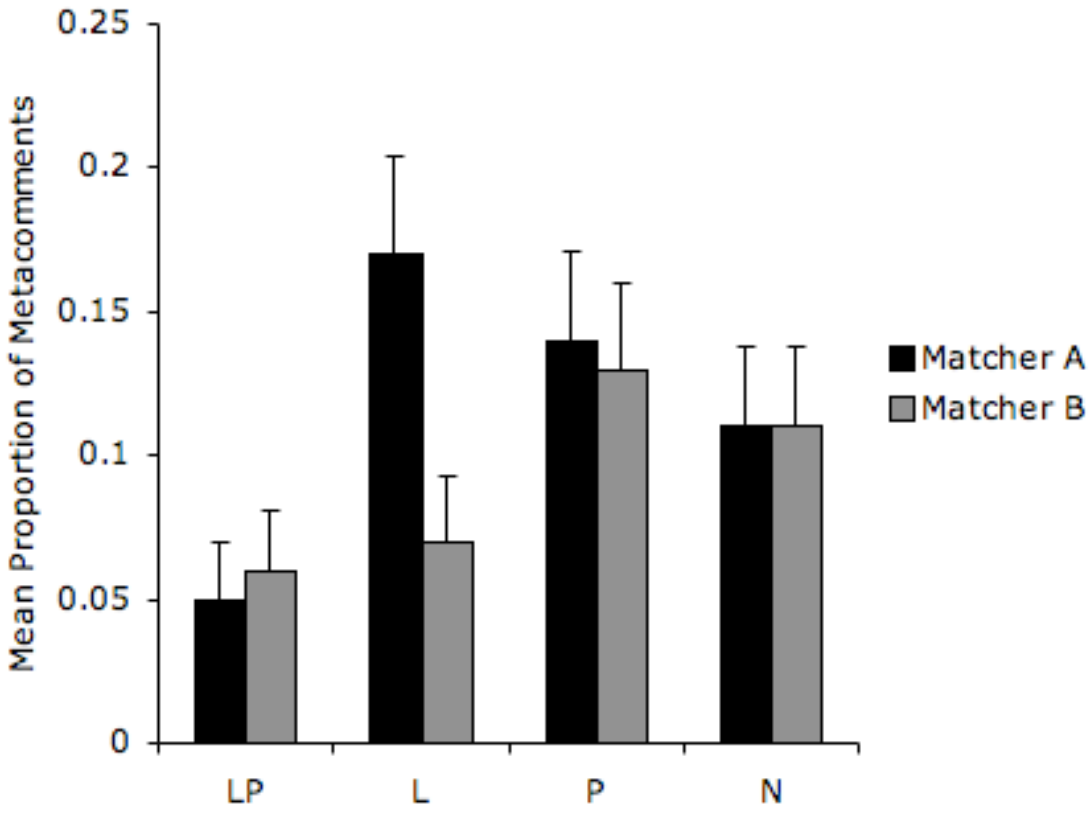
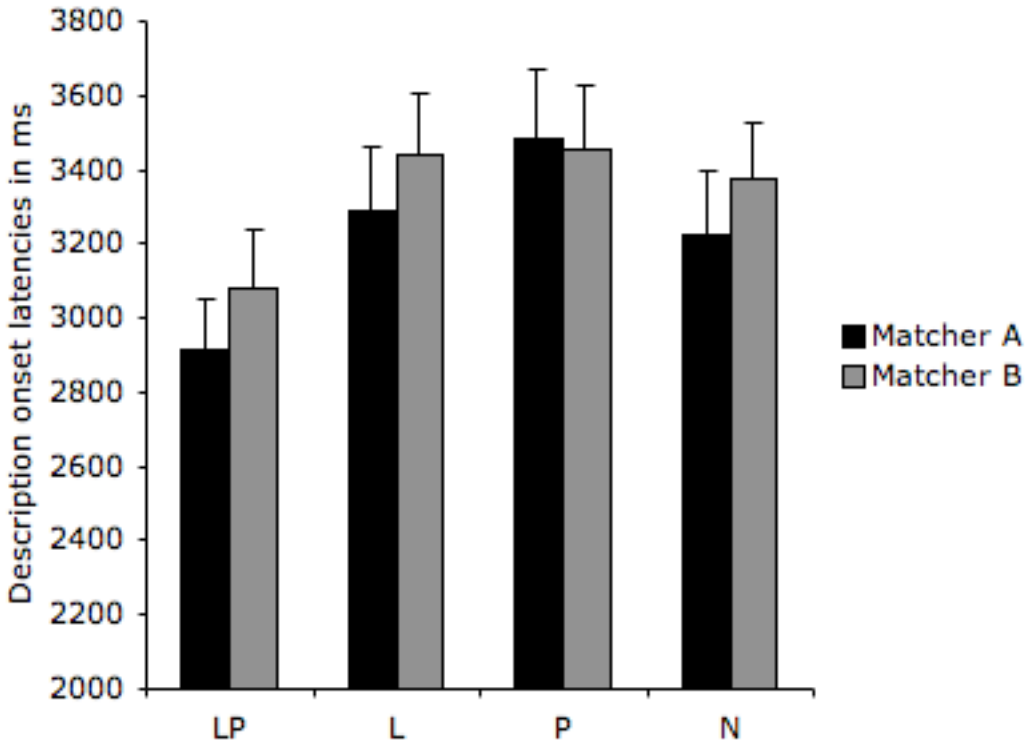


Figure 8
Experiment 1, Mean Description Onset Latencies to Produce the First Content Word for the Description in Phase 2, according to the Co-Presence Conditions of Phase 1 and Matcher Identity. Bars represent standard errors.



Appendix F
Tables 1 through 5

Table 1
Distribution of Items and Conditions of Co-Presence according to the Design of Experiments 1 and 2. Jokers were placeholder cards. (Item numbers refer to the identity of cards, not to the order in which they are matched.)

Phase 1: Establishing Shared Information

Distribution of items

	Round 1 with Matcher A	Round 2 with Matcher B
Director has:	item 1, item 2, item 3, item 4	item 5, item 6, item 7, item 8
Matcher has:	item 1, item 2, item 5, item 6 & 2 jokers	item 3, item 4, item 7, item 8 & 2 jokers

Information status of items

Subset	Round 1 with Matcher A	Round 2 with Matcher B
1. *items 1 & 2*:	shared linguistically and physically	not shared
2. *items 3 & 4:*	shared linguistically only	shared physically only
3. *items 5 & 6:*	shared physically only	shared linguistically only
4. *items 7 & 8*:	not shared	shared linguistically and physically

Phase 2: Assessing Shared information

Director matches all cards with each Matcher in Rounds 3 and 4
The order of Matchers is counterbalanced (ABAB or ABBA).

Table 2
Experiment 1, Means and standard deviations for the total number of words, total number of idea units, proportion of items with hedges, proportion of items with metacomments, number of turns, errors, and proportions of Matchers' contribution types (Acceptance, Clarification request, and Expansion request) involved in describing tangrams that were shared linguistically and physically or only linguistically in Phase 1.

	Linguistically and Physically	Only Linguistically
Number of words		
M	36.20	62.19
SD	29.63	46.42
Number of idea units		
M	8.85	15.13
SD	6.46	11.46
Proportion of descriptions with hedges		
M	.46	.59
SD	.50	.49
Proportion of descriptions with metacomments		
M	.25	.39
SD	.43	.49
Number of turns		
M	4.86	7.31
SD	3.85	6.07
Number of Errors		
M	.05	.20
SD	.28	.48
Proportion of Matcher's Contribution: Acceptance		
M	.50	.38
SD	.50	.48
Proportion of Matcher's Contribution: Clarification request		
M	.06	.15
SD	.24	.36
Proportion of Matcher's Contribution: Expansion request		
M	.29	.41
SD	.45	.49

Table 3
Distribution of triplets culled from Experiment 2 across the four co-presence and Matcher conditions and the two Matcher orders.

	ABAB	ABBA	Total
1. LP for MA; N for MB	9	11	20
2: L for MA; P for MB	12	10	22
3. P for MA; L for MB	11	11	22
4. N for MA; LP for MB	9	11	20
Total	41	43	84

Table 4
Experiments 1 and 2, Proportions of hits and false alarms, and d' values according to the co-presence conditions of Phase 1.

	Hits	False alarms	d'
Experiment 1			
Shared Linguistically and Physically	.86	.57	.90
Shared Linguistically only	.38	.17	.65
Shared Physically only	.18	.04	.84
Completely New	.13	.04	.62
Experiment 2			
Shared Linguistically and Physically	.79	.54	.71
Shared Linguistically only	.19	.10	.40
Shared Physically only	.20	.03	1.04
Completely New	.12	.04	.58

Table 5
Experiment 3, means and standard deviations for ratings of the relative clarity of lexically identical expressions according to the information status of the mention, Matcher order, and Matcher identity.

	Mention 1	Mention 2: Linguistically & Physically	Mention 2: Linguistically only	Mention 2: Physically only	Mention 2: New
Order ABAB					
Matcher A					
M	3.73	3.53	3.70	3.70	3.81
SD	1.14	1.09	1.20	1.16	1.05
Matcher B					
M	3.90	3.64	3.70	3.81	3.28
SD	1.05	1.11	1.14	1.04	1.22
Order ABBA					
Matcher A					
M	3.66	3.42	3.36	3.45	3.36
SD	1.13	1.20	1.27	1.19	1.22
Matcher B					
M	3.70	3.37	3.46	3.69	3.46
SD	1.14	1.25	1.22	1.06	1.19