Impact of Working Memory Span on Referring in Conversation

A Dissertation Presented

by

Calion Barry Lockridge Jr.

to

The Graduate School

in Partial fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Experimental Psychology

Stony Brook University

August 2007

Stony Brook University

The Graduate School

Calion Barry Lockridge, Jr.

We, the dissertation committee for the above candidate for the

Doctor of Philosophy degree,

hereby recommend acceptance of this dissertation.

Dissertation Advisor
Susan Brennan, Associate Professor, Psychology Department

Chairperson of the Defense
Dr. Richard Gerrig, Professor, Psychology Department

Dr. Antonio Freitas, Assistant Professor, Psychology Department

Dr. Marie Huffman, Associate Professor, Linguistics Department, Stony Brook University

This dissertation is accepted by the Graduate School

Lawrence Martin
Dean of the Graduate School

Abstract of the Dissertation

Impact of Working Memory Span on Referring in Conversation

by

Calion Barry Lockridge, Jr.

Doctor of Philosophy

in

Experimental Psychology

Stony Brook University

2007

This dissertation project investigated the possibility that individual differences in working memory span can shape conversation, specifically, how two people set a joint perspective. A two-phase experiment was used to investigate this question. During Phase I, participants were administered a series of individual difference tests, consisting of one primary (Operation Span) and two other secondary working memory tests (NBack, CVLT), as well as a questionnaire (Interpersonal Reactivity Index) to determine their personal inclination to take their partner's perspective. After being identified by the primary WM measure as having a high or low working memory span, participants in Phase 2 were separated into pairs and assigned to one of four conditions consisting of different combinations of their individual working memory span and their role (director-matcher) in the communication task (e.g. High-High, High-Low, Low-High and Low-Low). Pairs were given 18 picture cards to match over 5 rounds. Pairs' prior knowledge about each card's perspective was varied by whether they had learned the Same, Different, or No perspectives beforehand. The results suggest a relationship between individual working memory span and perspective setting behavior, and also suggested that pairs' effort and accuracy when setting a joint perspective is highly influenced by prior knowledge.

for my family

Table of Contents

List of Figures

List of Tables

Acknowledgments

Throughout the doctoral process, Susan Brennan, has served as both mentor and advisor and I thank her for all of her assistance and guidance.

I greatly appreciate the many years of generous support from the W. Burghardt Turner Fellowship and those who ran and operated it during my tenure at Stony Brook: Olufemi Vaughan, Anthony Hurley, and Phyllis Brenna.

I thank the members of my dissertation committee: Susan Brennan, Richard Gerrig, Marie Huffman, and Antonio Frietas, for their helpful advice and encouragement. In particular, I thank Dr. Frietas for consenting to advise this project so near its end.

Every researcher needs good research assistants, and I had some of the best. Thank you Anny Cheng, Randy Stein, Elizabeth Cohen for the mundane tasks like making photocopies, to the more challenging ones of running test subjects.

One’s cohort can be invaluable and I would like to thank both Shawn Bediako and Justina Ohaeri-Ekeocha for being sources of guidance and encouragement.

This research was supported by the W. Burghardt Turner Fellowship, Ruth L. Kirschstein National Research Service Award, and the Visiting Scholar’s Program at Allegheny College.

Chapter 1. Introduction

This dissertation project aims to understand how an individual difference in basic cognitive resources – working memory span – affects language use during conversation. Typically, laboratory studies of conversation do not look at individual differences at all, but model generic speakers and addressees. However, individual differences are particularly interesting for at least two reasons. First, a very limited set of studies have addressed how adjustment or coordination processes work when two partners are in a conversation and do *not* share the same capabilities or characteristics (e.g., see Bortfeld & Brennan, 1997 for dyads consisting of native and non-native English speakers; Isaacs & Clark, 1987, for experts and novices; Schober 1998 for differences in spatial ability). Second, while the proposed project explores the effects of individual differences in normal, healthy speakers on referential communication, this work could also have application to populations whose working memory span is impaired (see, e.g., Caplan, 1992). Although individual differences in memory span have been studied in the context of reading (e.g. Just & Carpenter,1992; Perlmutter & MacDonald,1995), they have not been studied in face to face conversation.

It is likely that individual differences in working memory span would have a significant impact, since people in conversation are engaged in a remarkable sort of multi-tasking. Given the diversity of information relevant to a conversation, the degree of flexibility speakers and addressees have in producing and comprehending utterances should be influenced by working memory span. While engaged in conversation, people attempt to dynamically monitor and manage salient information and perform conversational subtasks more efficiently such as:

Given vs. New Information. People must integrate current information with information from previous speech, recognizing “given” information and using it to integrate new information into the discourse model. This isn’t a difficult job for interlocutors because they can recognize what is given and new from the ongoing conversation from available discourse cues (Garrod & Pickering, 2004). These cues can also be used to assist interlocutors to converge on a common term that can be reused later to refer to same thing, and this conversational behavior between interlocutors has been referred to as lexical entrainment (Bortfeld & Brennan, 1997; Brennan & Clark, 1996; Garrod & Anderson, 1987). Although not the same thing as entrainment, efficient management of given and new information can make the process of entrainment more efficient.

Task Related Goals. This information includes keeping in mind any goal(s) a speaker has with respect to a conversation, as well as how speakers and addressees set their criteria for meeting such goals. Goals may or may not be shared between two participants, and it may take time for a conflict in goals to become apparent over the course of the conversation (Wilkes-Gibbs, 1986). In addition, for a given task, one person may have a higher criterion for success than the other and thus they may need to re-negotiate this discrepancy in order to accomplish any joint task.

Discrepancies in perspective. People monitor and accommodate possible discrepancies in perspective with their conversational partners. Discrepancies can arise from differences in spatial perspective (Schober, 1995); roles, or the distribution of task information (Bard, Anderson, Sotillo, Aylett, & Doherty-Sneddon, 2000; Clark & Wilkes-Gibbs, 1986; Russell & Schober, 1999), and these discrepancies must be taken account of by people in conversation in order to communicate effectively.

Audience Design. A conversational subtask where people strive to take the needs of their conversational partners into account while comprehending and producing language. While speaking, a conversational partner can shape how their addressees plan what to say. For example, people might recount what happened at a college party differently to their parents than to their friends; and they might have a difficult time deciding what details to report if both types of addressees (parents and friends) were present in the same room.

Perspective Setting. Conversational subtask that refers to how people conceptualize an object in order to refer to it distinctively. How people conceptualize objects in the world that are the topic of conversation could be influenced by how much cognitive resources they have to accommodate the relevant features of the object, along with any additional salient information from the conversational situation.

The likelihood exists that working memory span may not be the only factor that could shape referential communication. In addition to individual cognitive differences such as working memory span, social traits possessed by conversationalists (ex. willingness to take a partner's perspective) could have an impact as well.

In my dissertation, I addressed these questions:

- To what extent does working memory span impact referential communication, perspective setting in particular, in both the speaker and addressee role?
- How does working memory span affect how people set a joint perspective when their prior knowledge about the topic of conversation varies?

1.1 The Enterprise

First, I will review what we know so far about language processing and working memory capacity. Then I will discuss how working memory capacity could impact the products (e.g. successful references, lexical entrainment, common ground) and coordination processes of conversation. Next, I will talk about what has been shown in the literature with respect to working memory span and referential communication. I will present the referential communication task (RCT) as an experimental paradigm to best study the impact of working memory span on conversation, followed by a description of its working memory requirements. Then I will discuss individual differences in working memory span in the context of two aspects of referential communication: audience design and perspective taking. Finally, I will present my experimental paradigm along with a description of my dependent measures, some specific predictions, and the primary comparisons of interest.

1.2. Working Memory Capacity and Language Use

Working memory is critical for all complex cognitive activities, playing a role in reasoning, learning, problem solving, and language use. Working memory is the temporary reserve where immediate and final products of moment-by-moment cognitive processing occur (Baddeley, 1994; Baddeley & Hitch, 1974; Just & Carpenter, 1992). Working memory has a limited capacity, many researchers (Baddeley, 1994; Miller, 1956) have attempted to measure its limitations and have discovered that its capacity varies with the experimental task (Schunn, Lovett, & Reder, 2001; Shah & Miyake, 1996), as well as a person's expertise with the stimuli (Gobet & Simon, 1996; Chase & Simon, 1973).

In the field of psycholinguistics, working memory capacity - in the form of memory span (Daneman & Carpenter, 1980) – has been used to investigate differences in reading comprehension among readers of different skill. Readers access and process information in a dynamic fashion, often times using information read previously to comprehend subsequent material. In particular, findings in the reading comprehension literature show that memory span can predict individual differences in how people use context to resolve syntactic ambiguities (Miyake, Carpenter, & Just 1994; Perlmutter & MacDonald, 1995), understand complex grammatical structures (Just & Carpenter, 1992), and update current representations of a text (Palladino, Cornoldi, De Beni, & Pazzaglia, 2001). However, there are no studies in this literature that examine how individual differences in working memory span impact spontaneous spoken language. In the next section, I will describe working memory capacity and its possible impact on conversation using two theoretical frameworks that refer to aspects of the social context and structure of conversation common ground and the contribution model.

1.3. Working Memory Capacity and its Possible Impact on Conversation

To examine how individual differences in working memory span should impact conversation, I considered the theoretical perspectives of common ground and the collaboration model. Both the common ground and collaboration model are part of a larger theory of conversational interaction called the *collaborative view* that states that interlocutors use the processes and products of their moment by moment collaborations during conversation to ensure their mutual understanding (Clark & Wilkes-Gibbs, 1986; Schober & Clark, 1989). Common ground describes the sources of information in conversation that serve to shape the products of people's efforts to establish mutual knowledge. The collaborative view considers the process and structure of how people coordinate their conversational contributions to establish mutual knowledge. I will briefly define common ground, and then present two proposals that describe how common ground is encoded in memory during conversation. Then I will illustrate ways individual differences in working memory span can influence how people manage common ground during conversation. Finally, I will use two conversational examples to illustrate how individual differences in working memory span could shape the processes involved with coordinating contributions to an ongoing discourse.

1.3.1. Working Memory Capacity and Common Ground

Common ground consists of information people in conversation believe is mutually known, and may be inferred from three sources: (a) community co-membership, knowledge that people mutually believe they share based on membership in a given community; (b) physical co-presence, perceptual knowledge people share, based on being able to mutually hear or see each other or elements in their environment; and (c) linguistic co-presence, shared knowledge based on the ongoing discourse record (Clark and Marshall, 1981). According to Clark and colleagues, common ground is accrued and managed between participants through joint conversational interaction.

Clark and Wilkes-Gibbs (1986) suggest that the way common ground and memory could interact was that common ground may serve to flexibly influence domains of reference, moderating how interlocutors resolve referring expressions.

Clark and Wilkes-Gibbs (1986) also proposed their idea without delineating between working memory and any other type of memory in particular. For example, when two people are physically co-present with each other and the item that they are talking about, it would be unlikely that either partner would refer to that item without taking into account relevant information in the environment. A strong form of this view (attributed to Clark by Keysar, 1997 and colleagues) is that common ground inflexibly *restricts* how people establish domains of reference, and represents a special or modular kind of information in human memory (Keysar, 1997; Keysar, Barr, & Horton, 1998; Keysar, Barr, Balin, & Brauner, 2000).

Alternatively, Hanna and Tanenhaus (2003), Horton and Gerrig (2005a), and Lockridge and Brennan (2002) propose that common ground does not possesses any special status, but behaves just as any other information in memory. On that view, common ground doesn't completely restrict how people refer to objects but is a *probabilistic* constraint, interacting with the discourse structure and any additional contextual information in memory available at a particular time. Studies using eye tracking, a paradigm that allows a level of analysis that permits experimenters to make strong assumptions of the moment-by-moment properties of what a person is cognitively processing at any point in time, reveal that common ground interacts with the surrounding discourse to shape reference interpretation without special influence over and above any other information in memory (Hanna, 2001). In addition, the influence of common ground as a constraint of any kind may change with respect to the social context and discourse structure that surrounds a particular utterance, allowing people to converge on alternative interpretations. I propose that working memory capacity may have a measurable impact on how language processing is shaped by the mutual knowledge represented in common ground as well as other aspects of discourse context.

1.3.2. Working Memory Capacity and Coordination

It makes sense not to just think about the mutual knowledge people derive from their social setting, but how they coordinate its use. The contribution model, presented by Clark and Schaefer (1987) and revised by Cahn and Brennan (1999), is an attempt to formalize relevant aspects of this coordination. The contribution model states that every contribution to a conversation has two components: a presentation phase and an acceptance phase. Each utterance is itself a presentation; it does not become a contribution to an ongoing discourse until it is accepted. The acceptance phase is considered complete by the interlocutors when there is sufficient evidence that the addressee trusts that he understands what the speaker meant, and that the speaker recognizes this evidence.

There are no studies in the psycholinguistics literature that have examined how a person's working memory span impacts how they perform the presentation or acceptance phase of a contribution. One possibility is that people with a high working memory span could incorporate more information from common ground and adjust their contributions during either presentation or acceptance phases. On the other hand, a low working memory span may make it difficult for people to dynamically incorporate additional information from common ground to shape their conversational contributions. The following conversational exchange (Brennan & Lockridge, 2005) is between two people engaged in a referential communication task, where one person is the director (D) and is working with their partner (M) to arrange a group of 12 pictures in front of M to match D's array. D can see exactly where M is focusing her eyes on by virtue of a head mounted eye-tracker that shows M's precise eye fixations on the cards in front of her (the brackets, [], provide non-verbal information such as the position of M's eye cursor, which was visible to D):

> D: okay, number one is the basket that's long on both sides—
> you got it right there oop, go back to the right
> right there that's it, that's number one
>
> M: okay
>
> D: okay, number two is like—
> at the bottom, very small
> yup, you just had it right there that's two
>
> D: kay, number three has a handle going across
> it is—*oop, to the left yeah, right there your eyes are right on it
>
> *M:[*eye cursor is on the correct card*]*

M: this one?

D: that one? uh-huh number four is just very deep and hollow um, looks like the one straight up top

M: This one?

D: Yup

With her initial presentation, D begins by describing the first item at length (e.g. “number one is the basket that’s long on both sides”). D then refers to her partner’s eye fixation information in a separate sentence, seemingly as clarifying information included as an afterthought. When matching the second card D gives a shorter description, with a brief pause (to possibly check her partner’s focus of attention) and then uses her partner’s fixation information to direct her to the correct card. However, when matching card three, D interrupts herself in mid-sentence, “stopping on a dime” to inform M: that her “eyes are right on” the correct card. While describing card four, D appears to more closely integrate her descriptions of the item with evidence of their understanding via their moment-by-moment eye fixations. D appears to smoothly manage the two cognitive tasks used in her referring strategy, the first one where she describes the item, and the second where she checks for visual evidence of her partner’s understanding.

In contrast, consider the next conversation between a different pair (also designated D and M) where D has difficulty using his partner’s eye fixation in concert with describing the pictures:

D: The first one is the poodle.

M: [*picks up correct card and places it on display board*].

D: The second one is the dog with the side view, looking straight.

M: [*picks up incorrect card and places it on the display board*]

D: no, th-

M: um..

D: there's a picture from the side-the dog's looking straight forward.
its black, yeah, that one

M: [*grabs card*]

D: no…no, no, no, not that one

M: which one? - [*his hand briefly hovers over two cards, one of them the correct one*]

D: it's black – [*M:'s eye cursor glides between the same two cards, one of them being the correct one*]

M: this one? - [he points directly at the correct card and then picks it up and holds it in front of the head mounted eye camera for D to see].

D: yeah,

D: The third one's the husky.

Although the working memory spans of these people are unknown, both examples illustrate how working memory span *might* affect conversational strategies and contributions. A higher working memory span would allow a person to accommodate the increased load on working memory that results from trying to incorporate evidence of their partner's understanding while simultaneously making a conversational contribution. The working memory load would be high especially in situations where a speaker's evidence of their partner's understanding comes from non-static, rapidly changing sources (e.g. moment by moment eye fixations). In the first example, D demonstrates the impact of a high working memory capacity when she interrupts her contribution and immediately informs her partner that her "eyes were right on" the correct item. This is in contrast with the second example, where D has a more difficult time crafting his contribution, simultaneously monitoring the quickly changing evidence of his partner's understanding, and interrupting his speech in response. Although that project did not test individual differences in ability to combine channels in mediated communication, such differences in coordination ability might be related to individual differences in working memory span.

In the next section, I will take a look how individual differences in working memory span may impact referential communication by examining perspective setting and audience design. I will review the use of the referential communication task (RCT) as an experimental paradigm to explore the possible relationship between memory span and face to face conversation along with a description of the working memory requirements of the RCT.

1.4 The Referential Communication Task

The Referential Communication Task (RCT) is one of the most frequently used methods for studying language use in conversation. It can be traced back to ideas of Piaget in the 1920's (see, e.g., Schober & Brennan, 2003), and its current use is as an experimental paradigm for studying referring in dialogue, beginning in the 1960's (e.g., Krauss & Glucksberg, 1969; Krauss & Weinheimer, 1966). The RCT involves two people, one with information that the other needs, and they must collaborate in order to accomplish a joint goal. The person who knows the garget arrangement or location has often been called the director (Clark & Wilkes-Gibbs, 1986), or expert (Isaacs & Clark, 1987) and the other person has been called the matcher and novice. The actual information that both parties share varies by task. In some versions of the task, the director has to instruct the matcher to select an object out of a set of similar objects, and in others, direct their matcher along a predetermined route.

Among the advantages of this type of task is its ability to offer an array of potentially objective measures. For instance, a person’s intentions are less ambiguous because they are constrained by the task, and the level of comprehension can be measured using task performance. At a gross level, measures can consist of the degree of success or failure at the task and the amount of time people take to complete it. At an intermediate level, measures can include how a speaker chooses to organize discourse units. One reasonable assumption is that when a linguistic constituent is available in speakers' working memory, it is available to be selected and produced within the utterance (Ferreira & Dell, 2000) and that elements that are available early probably appeared in an early position an utterance, (see Bock, 1986; Brown & Dell, 1987; Levelt, 1989; Lockridge & Brennan, 2002). At a fine level, measures include behaviors that are thought to be closely coupled with moment by moment cognitive processing, such as the coordinated hand movements associated with reaching for an object, in concert with an accompanying spoken direction (Trueswell, Sekerina, Hill, & Logrip, 1999), making computer mouse movements over a display (Brennan, 1990; 2005), and eye movements (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995).

An additional advantage of the RCT is that its basic arrangement affords experimenter control over variations in the types of material to be discussed, and the set of stimuli targets the matcher must choose from. Experimenters can also control how often people must refer to items, and whether they switched roles with their partner to examine hypotheses based on the specific conversational role (speaker or addressee) that people are assigned during the RCT. Also, they can control whether either person is another experimental participant or a confederate. Most important, the RCT affords experimenters control over the moment-by-moment working memory load the situation places upon speakers and addressees during conversation.

During the RCT, a possible way to easily vary the working memory load is to manipulate the lexicalization of the items that are the topic of conversation (Gergle, Kraut, & Fussell, 2004; Kraut, Gergle, & Fussell, 2002). An item is lexicalized if it is something that has a common and available name in memory, such as the label “Doberman” for a picture of a Doberman pinscher. Alternatively, an object is not lexicalized when it does not have a familiar label, which is the case with geometric figures called tangrams, for which people must generate novel perspectives and referring expressions. Also, lexicalization can pertain to an object’s typicality; even within a familiar category (e.g. birds), people are generally more likely to have lexicalized labels for a picture of a familiar bird (like a robin) than for an unfamiliar bird (like an osprey). Lexicalized items, such as pictures of common things like dogs, should place less of a burden on working memory and be easier for directors and matchers to coordinate their references for, as opposed to non-lexicalized items that are harder to describe and match (Gergle, Kraut, & Fussell, 2004; Kraut, Gergle, & Fussell, 2002).

Studies of language use based on the RCT have sometimes focused on hypotheses related to either production or comprehension. For example, research that focuses its analysis on the behavior of the director in the RCT is believed to reveal information based on aspects of language production, such as audience design (e.g. Bortfeld & Brennan, 1997; Clark & Wilkes-Gibbs 1986; Horton & Gerrig, 2002; Horton & Keysar, 1996; Isaacs & Clark 1987). On the other hand, research that focuses its analysis on the behavior of the matcher is believed to reveal aspects of spoken language comprehension, such as how addressees alter their behavior with respect to how speakers violate conceptual pacts (Metzing & Brennan, 2003).

Although comprehension and production can be considered separately, in reality these processes are closely integrated. Rarely during everyday conversation or during the referential communication task do people comprehend or produce language exclusively with one process uninfluenced by the other. Language is produced and comprehended in parallel, with one process perhaps dominant, depending on whether an individual is talking or listening. For this reason, the referential communication task is a good method for studying spoken language production and comprehension in parallel. In the following section, I will discuss the primary working memory requirements of the referential communication task.

1.4.1 The Working Memory Requirements of the Referential Communication Task

Several aspects of the referential communication task can be straightforwardly linked to its primary memory requirements. Those aspects are the item array (the kinds, numbers, and distinctiveness of objects that are being discussed), experience with the particular item set (due to matching the same items over repeated trials), the visual environment (what either person can and cannot see during the task), and role (whether or not a person is acting as the director or matcher)

Item Array. The item array can be described as the objects that form the topic of conversation during the referential communication task. The items can place a load on working memory with respect to their number and the degree that they are lexicalized. The number of items that people have to keep track of during conversation can make it difficult to form perspectives when the number of items increases. When items are non-lexicalized, it is harder to form a joint perspective (Gergle, Kraut, & Fussell, 2004; Kraut, Gergle, & Fussell, 2002). One rationale for that difficulty is that more cognitive effort is expended (reducing cognitive resources for other activities) such as creating new representations "on the fly".

Experience. The experience factor relates to the working memory load people encounter based on accumulating information from the experience of creating joint perspectives with their partner. For example, although a referential communication task becomes easier when directors and matchers repeatedly match the same objects, it can become an increasing load on working memory when directors have to keep track of their experiences with different items or subsets of items. This *experiential* information can pertain to metacognitive monitoring processes done by directors to keep track of whether they are their matcher has experienced a perspective or not. Another source of experiential information is the amount of time and effort a director and matcher might spend trying to converge on a joint perspective for an item. For instance, if a director notices that they are spending a disproportionate amount of time and effort matching an item or subset of items, they could use this information to potentially enhance their performance in the future. They could use this information as cues to anticipate when they should expend more effort or develop alternative referring expressions.

Visual Environment. The visual environment of the referential communication task as a memory requirement relates to what is visible either the director or matcher can or cannot see during the task. Whether or not directors and matchers have access to visual information relevant to the item array and each other can impact how well they can ground their conversation. For example, if two people have visual access to each other and what they are talking about, grounding their conversation might be easy and efficient because they could visually infer more information about how well their partner understands what they have said. Alternatively, grounding might be more difficult if they had to rely on verbal information exclusively, such as over the phone (Brennan & Lockridge, 2006). The RCT would increase its demands on working memory by reducing the amount of visual information available for grounding.

Role. It is important to note that in ordinary conversation, people are rapidly alternating between the roles of speaker and addressee. The implication is that the processes of speaking and hearing, though often studied in isolation, are likely to be influenced by each other when they take place in an interactive setting (Pickering & Garrod, 2004). People in conversation produce and comprehend language in parallel, with one process perhaps somewhat dominant for the moment depending on whether they are spending more time talking or listening.

The working memory requirements of the referential communication task would have different implications for people in the experimenter-designated roles of director and matcher. Directors typically produce more speech than matchers during the RCT, and since much of the initiative for coordinating and "stage managing" falls on the director's shoulders, it could be said that their role is more working memory intensive than the matchers. A possibility is that the RCT is less difficult for matchers when they have the same items as their partner, placing them in a forced choice situation. They don't have to figure out whether something is present or not, they just have to engage in a recognition task that is easier and less taxing on working memory.

Also contributing to the increased load on their working memory is the fact that directors have to craft verbal descriptions that take into account additional sources of information such as the feedback received from their matchers that serves as evidence of how well they understand. Evidence from the matcher doesn't have to be verbal; it can also be visual, such as situations where the matcher places an item in its target position in the item array when the Director can see what they are doing. Different kinds of feedback can make directors and matchers engage in alternate strategies to refer to the items in the item array, and each strategy can vary the load performing the RCT takes on working memory. For example, when directors and matchers have to rely exclusively on verbal evidence of understanding, they tend to use more words to describe the target items (Brennan, 1990, 2005, Brennan & Lockridge, 2005). However, if they have more visual evidence available, then their descriptions shorten because some of the information contained within the longer descriptions can be inferred visually.

For instance, let's say a director is having trouble getting a matcher to understand a description of a particularly difficult item. If the director can see the where the matcher is looking (as in the first of the previous two examples), then they might use the matcher's focus of attention (Doherty-Sneddon, Anderson, O'Malley, Langton, Garrod, & Bruce, 1997), or the matcher's actions with respect to their current orientation to the objects in the item array (e.g., over which object their hand is poised) to determine their current level of understanding (Clark & Krych, 2004). With such visual evidence, directors and matchers can infer each other's mutual knowledge or understanding moment-by-moment, they don't need to use longer verbal descriptions that could increase their working memory load during conversation. When two people both use the same shortened terms in referring expressions addressed to each other, it can be assumed that a shared joint perspective has already been formed, and less working memory resources should be required to reuse that term in the future. Alternatively, longer terms means a mental representation hasn't been formed and usage of those terms could increase the working memory load.

The working memory requirements for matchers, on the other hand, stem from their task of trying to understand the perspectives implicit in directors' expression and map the features of a perspective onto their representation of the target item in working memory. They have to match features in accordance with any additional memory cues from the other parts of the RCT that are primary working memory requirements. matchers maintain their representation in working memory as they scan the item array for the target item while considering whether or not they need to request more information to improve their understanding of a referring expression. Although matchers are far from passive and silent during the referential communication task, presumably they spend most of their time comprehending spoken discourse and planning instrumental actions. In the next section, I will discuss how having a high or low working memory span should impact how directors and matchers do several cognitive and communicative subtasks during the RCT.

1.4.2. How Individual Differences in Working Memory Span may affect Performance on the Referential Communication Task

Several primary cognitive subtasks in conversation may be influenced by working memory span, such as: a) distinguishing objects in an array of similar objects, b) mapping referring expressions onto objects, c) mapping objects onto referring expressions, and d) repairing misunderstandings.

Evaluating Differences Among Objects. For matchers, possessing a high memory span might allow the use of more effective searching and grouping strategies while searching for objects in the item array, thereby allowing them to better evaluate the differences among the objects. While scanning the item array, matchers could encounter target items in the presence of similar competitors that could share relevant features. With more resources available, high span matchers may be able to rapidly notice and categorize important distinguishing features that may not have appeared in the director's initial referring expression. Additional features high span matchers should be able to encode could include an item's orientation, or they may be better able to propose additional salient features to enhance the director's description. Later, when a director re-refers to a target object in the item array, a high span matcher should have an easier time selecting it because of the additional self-generated memory cue. Low span matchers should be less effective than their higher span counterparts at encoding salient features of target items not contained in the director's referring expressions.

As for directors, a higher working memory span may make them better at noticing relevant distinguishing features among the objects in the item array. For example, let's say that a director is describing a target item that shares many features with competitors in the item array. With a higher memory span, the director could encode more of the features of the target that would better distinguish it from its competitors, enabling them to craft a referring expression that encodes the relevant differences between items.

Mapping Referring Expressions onto Objects. Since a matcher with a high working memory span should be able to maintain more potential referents or features in working memory, they could better at encoding and mapping the different features of the objects in the item array to the director's referring expressions. At the moment a matcher hears the director's referring expression, they begin to map the expression onto relevant features of the most likely target(s) in the item array. With the working memory resources available to encode more of the relevant qualities of the director's referring expression into their mental representation of the target item, high span matchers may craft more distinct representations of the target and the other objects in the item array. Alternatively, a director's high working memory span may be able to maintain a more distinct representation of an item and craft referring expressions that could help matchers map their description onto the intended target item.

Repairing Misunderstandings. Directors with greater working memory spans have extra resources that might allow them to quickly and efficiently handle misunderstandings. If a matcher misunderstands them and selects the wrong item or nothing at all, directors might try to repair the misunderstanding through creating and proposing alternate perspectives, or they might also use features of an item the pair considered in an earlier round and prompting the matcher with those. Directors with high spans may also show greater flexibility by setting aside referring expressions when they prove to be ineffective. Matchers with high working memory spans may be more adept at generating counter-proposals if they judge a director's proposed perspective to be unclear. In the next section, I will further discuss the topic of individual differences and working memory span in the context of two basic conversation sub-tasks, audience design and perspective setting.

1.5. How Individual Differences in Working Memory Span may Affect Perspective Setting and Audience Design

What information there is in the psycholinguistics literature on individual differences and referential communication does not refer to memory span directly. When people in conversation refer to objects, they collaborate to set a particular *perspective* using terms that they will use to refer and re-refer to a particular object in the future (Schober & Clark, 1989;Wilkes-Gibbs & Clark, 1992).

People with high working memory spans could be better at flexibly managing the perspectives of either themselves or their partner(s), especially in situations where a prior perspective or a composite perspective would be beneficial.

Stellmann and Brennan (1993) investigated how flexible people were in setting perspectives during their experiences with two different partners. The question was, will people who have already established a perspective on an object with one partner, rely mainly on that perspective, flexibly adjust to the new partner's perspective, or create an entirely new joint perspective when paired with a new partner? Put another way, to what extent did they try to rely on (or be unable to leave behind) the previously-created joint perspective when speaking with the new partner? To answer this question, Stellmann and Brennan used a referential communication task with a set of 12 tangrams as the objects to be matched. Pairs of participants performed the matching task for twelve rounds, switching to a new partner for rounds 5-8, and returning to their original partner for the final four rounds. The first partner switch was staged to detect how flexible people would be in setting a joint perspective with a new partner when they had already set one with an old partner. The second switch was to examine whether reuniting with the old partner would cause people to return to their original (e.g. privately held) perspectives. The findings were that people used flexible perspective setting strategies that were based more on collaboration with their partner (supporting a partner-driven model) rather than relying only on their prior experience (Stellmann & Brennan, 1993). With new partners (in rounds 5-8), people often began by proposing the prior perspectives that had been successfully used with previous partners, and waited for their new partners to accept them. If their new partners didn't accept the proposed "old" perspective, then the new partner's perspective was adopted. Although they were all quite familiar with the objects (and they knew their partners were too), people took as much time (and as many words) to ground perspectives with these new partners as with their original partners.

While these results supported the idea that people are generally flexible with regard to perspective setting, the fact that a variety of different strategies were used suggests that individual differences in working memory span could have had an impact on the results. People with high memory spans may be able to better create composite or hybrid perspectives that consist of material from their own and their partner's perspectives. Stellmann and Brennan (1993) counted composite perspectives with major elements of previously used perspectives established in earlier trials. The ability to create hybrid perspectives and lexibly adjust to a partner's perspective should benefit people in both director and matcher roles.

When people use and reuse jointly set perspectives over time, they engage in entrainment (Bortfeld & Brennan, 1997; Brennan & Clark, 1996). This conversational behavior provides evidence of audience design. In situations where people must interact with more than one person, they must be flexible in the number of different perspectives they may encounter with each partner. Matching sets of objects with different matchers is a challenging task that can influence when and how directors engage in audience design (Horton & Gerrig, 2002). Matchers who interact with multiple partners also keep track of partner-relevant information that affects interpretation of referring expressions (Metzing & Brennan, 2003). During conversation, there may be memory cues generated as products of interaction that can enhance a director's ability to differentiate perspectives and use them appropriately with different partners.

High span directors might more reliably track success and failure in relation to their descriptions to a particular subset of items, allowing them to make relevant adjustments to the labels they remember and entrain on. For example, imagine that a director with a high working memory span has a hard time establishing a perspective on an object that is acceptable to the matcher. Having a high span may afford the director the ability to come up with several alternate perspectives, as well as to remember the relative difficulty of coming up with an acceptable perspective. If they encounter that same item later, the director may be better able to propose a perspective that the matcher finds acceptable.

Another way that directors can engage in audience design is to incorporate alternative perspectives from their matchers into their descriptions, creating a composite, hybrid perspective. I predict that directors with higher working memory spans should be better at incorporating relevant information from their matcher into their descriptions because they have the additional working memory resources to do so. These additional resources allow them to monitor the success and failure of their own descriptions and give them the flexibility to monitor their matchers' feedback and counterproposals to decide what information is salient and useful for integration into later perspectives. There is no reason to expect that directors with higher working memory spans would entrain differently from people with lower memory spans. Directors with higher spans may entrain in a way that reveals a greater ability to monitor and manage more information moment-by-moment.

Not only does a higher memory span enhance a director's skill in managing perspectives while engaging in audience design, it may enhance a matcher's ability to help the director. Matchers can assist directors when they are trying to construct a joint perspective by supplying informative feedback. Different types of feedback from a matcher can vary in how the director is prompted to engage in audience design. For example, verbal feedback can range from a simple acceptance, like "Okay" to a more elaborate response, such as when a matcher interrupts the director and asks for more information.

Higher working memory resources could also allow both directors and matchers to influence the balance of the director's effort, the percentage of words the director uses over the total amount uttered by both people with respect to creating a joint perspective. In the next section, I will summarize the prior sections and give a brief overview of the experimental paradigm.

1.6 Summary

The purpose of this introduction was to illustrate how individual differences in working memory span could shape face to face conversation, just as such differences have been demonstrated to shape other complex cognitive activities. The possibility that a person's working memory span could shape their conversational behavior hasn't been explored in the psycholinguistics literature for either spoken language production or comprehension. In fact, face to face conversation presents people with robust threads of information that arrive rapidly (e.g. information about the task, information from their partner, etc). A high working memory capacity on the part of one or both interlocutors might provide interlocutors with the ability to increase the chances that their conversation is efficient, as well as to better manage the products (i.e. mutual knowledge) and the processes of conversational coordination.

To empirically test hypotheses related to working memory span and spoken language use, the Referential Communication Task (RCT) appears to be the best available methodology. Not only does it allow examining spoken language production and comprehension in parallel, it allows controlling working memory requirements (e.g. item array, experience, visual environment, and role). Together these features of the RCT task contribute to the working memory load people experience. It is therefore important to understand and control the working memory requirements for participants.

With respect to referential communication, there are several primary cognitive subtasks that could prove difficult for people with low working memory spans to do. These cognitive subtasks are: efficiently distinguishing objects in the item array and applying this knowledge to generating referring expressions for what is the current referent (primarily for speakers), unpacking what is said and mapping it to a unique referent in the array (primarily for addressees), and repairing misunderstandings (for both speakers and addressees). How people do these subtasks during referential communication may depend on individual differences in their working memory span.

In the context of referential communication, specifically perspective setting, individual differences in working memory span can play a large role. Interlocutors with higher working memory spans should be able to flexibly construct, recall, and propose alternative perspectives, especially in situations where using a prior or a composite perspective could be beneficial. A high span could also allow interlocutors to dynamically integrate their specific experiences setting a joint perspective with particular items and partners into their referring expressions. In the next chapter, I will describe the overall experimental design in detail, along with my predictions.

Chapter 2. Method

2.1 Overall Design

To examine the question of how individual differences in working memory span may shape referential communication – in particular perspective setting – I conducted an experiment with two phases. Phase I used several experimental tasks to screen participants for their working memory capacity (Operation Span Task, NBack, CVLT) and their orientation toward taking their partner's perspective (IRI questionnaire). The primary measure for working memory span was the Operation Span (OSpan) task. The OSpan task is the most widely used task by researchers to investigate how people's working memory span can explain individual differences in performance on complex cognitive tasks, such as mathematic and verbal problem solving, strategy adaptivity, and fluid intelligence (Engle, Tuholski, Laughlin, Conway, 1999; Schunn & Reder, 1998; Staver & Jacks, 1988). Participants' performance on the OSpan task determined whether I categorized them as having a high or low working memory span.

In Phase II, pairs of participants did a variant of the referential communication task used in Wilkes-Gibbs and Kim (1991), during which they collaborated and arranged 18 3 x 5 inch picture cards of ambiguous geometric objects called tangrams (while separated by a visual barrier) on the matcher's display board. They did this for a total of 5 rounds. Participants were also assigned to be either the director or matcher (these roles were kept constant during the five matching rounds of the referential communication task). Each pair of participants was assigned to one of four conditions (between subjects):

(1) H_D-H_M: High span director / High span matcher
(2) H_D-L_M: High span director / Low span matcher
(3) L_D-H_M: Low span director / High span matcher
(4) L_D-L_M: Low span director / Low span matcher

This Wilkes-Gibbs and Kim (1991) version of the RCT allows control over what prior perspective participants held privately before they attempted to create a joint one. This is important because to measure how flexible a person was with respect to accommodating to their partner's perspective, we needed an idea of what privately held perspective they might have had to abandon or adjust.

This paradigm also allowed me to vary whether pairs' prior perspectives were aligned or not. I varied pairs' prior perspectives because during conversation it is rare that people find their own perspective and their partner's are consistently the same, different, or nonexistent. In fact, during conversation people are more likely to form joint perspectives while their own prior perspectives can be a combination of all three potential knowledge situations. The 18 tangrams that participants matched during the Phase II RCT were divided equally among three experimental conditions related to their prior perspective situation (within subjects): (a) *Same prior perspective*, (b) *Different prior perspective*, and (c) *No prior perspective*.

Both Phases of my experimental design will allow me to answer two primary questions about working memory span and referential communication; focusing on perspective setting:

- How does an individuals' working memory span influence their perspective setting behavior? Do high-span individuals demonstrate more flexibility, and are they more efficient?
- Do different prior perspectives interact with working memory span to shape how people set a joint perspective?

2.2. Overall Procedure

In Phase I, after they gave their informed consent, participants were given a battery of tests during the Phase I experimental session that lasted one hour. Based on their OSPAN score, participants were then assigned to the working memory span pairs and then contacted and scheduled to return for Phase II.

2.3 Participants

One hundred sixty three undergraduates who were native speakers of English participated in Phase I. Of these, 81 were selected and recruited to participate in Phase II as 41 pairs. Participants were matched with a partner and invited back to participate in Phase II based on their score on the Operation Span (OSpan) task in Phase I. They received either research participation credit or a $10 honorarium for each phase they participated in.

2.4 Materials

Stimuli. The 12 tangrams on the picture cards used by the matcher in Phase II were printed on 3 by 5 inch laminated index cards. In Phase II, the tangram pictures in the notebooks used by the participants during the study task and the tangrams in the notebook used by the director were printed in the center of an 8 1/2 by 11 inch page. The directors' tangram pictures were identical in size and orientation to the tangram pictures in the matchers' card set. Four tangrams appeared in the Different prior perspective condition, four appeared in the Same prior perspective condition, and the remaining four, in the No prior perspective condition.

Norming for Goodness of Labels. Sixteen tangrams and labels that encoded particular perspectives on them were normed for how "easy to see" the perspectives were, using a questionnaire. Twelve of the tangrams and pairs of alternative labels for them were taken from Wilkes-Gibbs and Kim (1991), and four tangrams were taken from a book (Elffers & Schuyt, 1997) and accompanying pairs of perspectives were created. Each tangram was presented individually in the center of an 8 ½ by 11 inch page with the following rating scale from one to five (see Figure 1).

Fifty people rated the tangrams and they were not participants in either Phase I or II of the experiment; they were recruited from a sophomore level English class and given a five-dollar honorarium for their participation.

Raters were instructed to circle the number on the scale that they felt best represented how well each label fit with its corresponding tangram. Tangram-label combinations were presented one to a page. Each person rated both labels for each tangram, but in different halves of the rating task, to prevent direct comparisons. The order that participants saw either perspective was counterbalanced using two versions of the questionnaires (e.g. half of the participants rated *barbell* then *dog bone* and the other half, *dog bone* then *barbell*). Participants were also instructed to refrain from looking at any of their earlier ratings. The rating task took about 10 minutes.

An ease-of-seeing rating was generated by taking the mean of a tangrams pair of ratings from both parts of the first and second version of the questionnaire (e.g. separate means for “barbell” and “dog bone”). The differences between ratings for these pairs of labels were used to select the tangrams for the prior perspective conditions. For example, tangrams chosen for the *Different prior perspective* condition had to have both labels rated at 3.0 or above (on a scale of 1 to 5) and the mean ratings for the two labels could not be significantly different statistically. This criterion yielded six tangrams (of the original 16) that had a pair of labels with equivalent ratings (e.g. *barbell* not significantly different from *dog bone*). For the pairs of perspectives with equivalent ease-of-seeing ratings, the mean of their pair of ratings ranged from 3.0 to 4.2 (e.g. dagger / necktie: both ease-of-seeing ratings = 3.0; barbell / dog bone: both ease-of-seeing = 4.2).

Tangrams with at least one of their pair of perspectives rated 3.5 or above were selected for the *Same prior perspective* condition, and the higher rated perspective label was used and the other discarded. This was because the Same prior perspective condition required that both participants learn only one perspective for the Phase II referential communication task. Tangrams chosen for the *No prior perspective* condition were chosen because neither of their labels fit the criteria for the same or different prior perspective condition. Appendix C shows pictures of all the tangrams used in each prior perspective condition.

2.5. Phase I Individual Difference Measures

Working memory (Operation Word Span Task (OSPAN); Lapointe & Engle, 1990). This task was used as the primary measure for working memory capacity, a task that requires participants to access their working memory in ways that don't rely solely on language (unlike another common WM measure, the reading span task, Daneman & Carpenter, 1980). Prior research has established the OSPAN task to be a reliable and valid marker of WM capacity (see Kane & Engle, 2003; for discussion). Participants performed this task on a computer.

Individual Tendency Toward Taking Partner's Perspective (IRI – Davis Interpersonal Reactivity Index). This is a questionnaire used to measure a person's general level of empathy towards others, with scales that refer to four individually discriminable personality variables Of these, only the *perspective taking* (a person's interest in taking or willingness to take another's perspective) scale was of interest. Although the perspective taking subscale does not measure participants' perspective taking *ability*, it served as a filter to allow me to rule out participants that were unusually resistant to taking their partner's perspective. The additional scales were: *fantasy* a person's tendency to transport themselves into the feelings and actions of fictitious characters in books, movies and television, *empathic concern* a person's tendency toward feelings of sympathy and concern for others, and *personal distress* a person's tendency toward feelings of personal anxiety and unease in tense situations (Davis, 1983). I administered IRI in its entirety so people would not focus on the perspective taking component.

2.5.1. Assignment to span conditions

Participants' performance on the OSpan task determined whether they would be placed in the high or low WM span category. Participants from the top and bottom quartiles of the distributions (see Figure 2) were designated as high and low span respectively, (High span Ospan score, M: 24.38, SD: 4.32, Low span Ospan score, M:9.12, SD: 2.81). Participants in the bottom quartile scored from 7–10 and were categorized as low span, and participants who scored from 20 and above were categorized as high span. I avoided using people who scored in the lowest of the bottom quartile, because of the possibility that people could "fake" a low OSpan score by not trying (note that it is not possible to fake a high score).

Since my distribution of OSpan scores was not complete before I started collecting data for Phase II, I initially used standard OSpan score designations from the psycholinguistics literature to place participants into high and low working memory span groups. I used an OSpan score of 20 as the approximate cutoff score for people designated as high working memory span and 10 for the low span cutoff because both scores have been used for individuals designated as high span and low span, respectively (Kane & Engle, 2003; Kane, Bleckley, Conway, & Engle, 2001; Tuholski, Engle, Baylis, 2001).

For the perspective taking measure from the IRI task, I found that high span directors did not significantly differ from low span directors, $F(1,30) = .270$, $p = .607$, n.s., nor did high span matchers differ from low span matchers, $F(1,30) = .436$, $p = .514$, n.s.(see Table 1).

Note that two additional tasks (NBack, California Verbal Learning Test – CVLT) were added to the design in response to NRSA reviewers of this dissertation project as a fellowship proposal. The NBack task (Braver, Cohen, Nystrom, Jonides, Smith, & Noll, 1997; Ragland, Turetsky, Gunning-Dixon, Turner, Schroeder, Chan, & Gurr, 2002) was used as a secondary working memory span measure to establish that the OSpan measure was not domain specific (see Kane, Conway, & Engle, 1999 for discussion). This task was also performed on a computer.

The CVLT (Delis, Kramer, Kaplan, & Ober, 1987; Norman, Evans, Walden-Miller, Heaton, 2000) is a neuropsychological test that can be used to assess an individual's long term memory for verbal information, along with subtests that measure short term memory for that same information. Also, this measure gives an indication of how well individuals transfer information from Working Memory into Long Term Memory, with respect to current theory that suggests working memory may include an additional temporary long term memory component (e.g. an episodic buffer; Baddeley, 2002). The CVLT is given in two parts, separated by a 20-minute delay

The order of the Phase I tasks were 1) OSpan, 2) IRI, 3) Part I of the CVLT, 4) NBack, 5) Part II of the CVLT. Participants from both the high and low working memory span groups performed poorly on the NBack task. They had a hard time identifying the target letters, and they consequently missed more than 50 percent of them. This problem made it difficult to try to correlate participants' OSpan scores with their performance on the NBack task. Participants' scores on the subscales of the CVLT also failed to correlate with their OSpan scores. It could have been the case that participants were less motivated by the time they did parts I and II of the CVLT because of the OSpan task and the NBack task. After performing both tasks, they could have been less motivated to expend their best effort Part II of the CVLT.

2.5.2. Procedure

Operation Word Span Task (OSPAN). Participants were seated in front of a computer screen where a Super Lab 5.0 program presented the stimuli. They were asked to solve a series of arithmetic operation-word strings presented with unrelated words to memorize. Operation-word strings were presented one at a time, and each block of operation-word strings ranged from two to six items in length. For instance, a block of four strings looked like:

IS (10 / 1) + 3 = 13 ? face

IS (10 x 2) + 2 = 21 ? jail

IS (9 x 3) - 2 = 25 ? point

IS (2 / 1) - 1 = 1 ? lamp

A participant read an arithmetic equation aloud as soon as it appeared on the screen. After reading the equation, the participant solved the equation, verifying out loud whether the given answer was correct and then immediately read the word aloud. Then they rapidly pressed the spacebar to proceed to the next arithmetic-word string. An experimenter was present just in case they needed additional prompting to not pause before reading the equation and to also not pause before moving on to the next operation word string. Participants would continue to solve operation word-strings until they encountered a series of eight question marks (????????) that served as a recall cue to signal when they were to only recall all of the words from the current block. For recall, participants were instructed to write as many of the words as they could remember on an answer sheet in the exact order that they had been presented. Each block size appeared in an unpredictable order that was fixed across participants to make the number of words to recall unknown until a participant encountered the recall cue.

Three blocks of each length (from two to six operation-word pairs) were presented, allowing scores to range from 0 to 60. The OSPAN score was the sum of the recalled words for only those blocks completely recalled in correct order. For instance, if a participant recalled only the two and three operation-word pair blocks completely and in correct order, their OSPAN score would be 15 (3 blocks of 2 operation-word pairs = 6, 3 blocks of 3 operation-word pairs = 9; 6 + 9 = 15 OSPAN score). In addition, participants who were to be considered for Phase II also had to have correctly verified the answers of at least 85 percent of the arithmetic strings.

Interpersonal Reactivity Index (IRI) Questionnaire. After the OSPAN task, the IRI instructions read aloud by the experimenter and participants then completed the IRI questionnaire. After giving instructions, before the experimenter left them alone they informed the participant to notify them when they finished the questionnaire.

2.6. Phase II – Referential Communication Task

2.6.1. Design

For matched working memory span pairs (H_D-H_M or L_D-L_M), each person in a pair was randomly assigned to the role of director or matcher. For mixed span pairs, depending on the condition each pair was in, either the high span or the low-span participant was assigned as either director or matcher. Pairs of participants came to the lab at the same time and went into separate rooms, where they first studied perspectives for 12 tangrams. For half of the tangrams, each pair of participants learned the same perspective, and for the other half both learned a different one with two equally easy to see perspectives (e.g., the director learned *dagger* while the matcher learned *necktie*; see Figure 3). To ensure learning, participants had a study phase where they were trained on the perspectives and labels they were supposed to learn.

In all, participants arranged 18 tangrams during the conversational task that corresponded to three experimental conditions – (1) 6 previously studied tangrams and labels for which they learned the same perspective as their partner during the study task – *Same prior perspective*, (2) 6 previously studied tangrams and labels where they studied a different perspective – *Different prior perspective*, and (3) 6 additional objects for which neither of them studied a perspective – *No prior perspective* (see Appendix C). The first three tangrams matched in each round were from the same prior perspective, different prior perspective, and no prior perspective conditions. I wanted participants to begin each round with tangrams where they had some prior information to avoid encouraging them to begin each round by focusing on the prior perspective condition. Tangrams chosen for the shared and the different learned perspective conditions were ordered for each round in Phase II so that both the easier and harder items based on ease-of-seeing ratings would have an opportunity to appear both at the beginning and end of the stimuli list during the experiment.

2.6.2 Procedure

Referential Communication Task (Instructions). Pairs of participants were informed that one of them would be the director, and the other would be the matcher (pairs were unaware that who would play what role was determined ahead of time). Then they were told that they would work together to match a series of 18 picture cards with tangrams on them for a total of five rounds. The director was shown the booklet containing the pre-ordered set of tangram pictures and told that their job was to describe them to the matcher. The matcher was instructed that their job would be to select the picture card from their set that they believed best matched the director's description among the eighteen that they had to choose from, and place it on the shelves on the poster board display in front of them (e.g. target area; see Figure 4).

Study Task. Next, before doing the matching task, they were led into separate rooms for the perspective study task. Once alone, they were told that for six of the twelve, they would learn the same perspective as their partner, and for the other half they would learn a different perspective.

I informed pairs of the different prior perspective condition of each tangram prior to the study task because I believed that if they knew that their prior perspectives might or might not vary with respect to each other's, they would be less likely to use a uniform strategy (e.g., to just assume their prior perspectives were either all shared with the partner, or all different). A review of the transcripts shows that pairs rarely bothered to focus on the perspective manipulation or use this kind of meta-task strategy; it may have been less costly to just start by proposing their own learned perspective and see if their partners agreed, as they could not jointly confirm any of their inferences about each tangram visually.

Each participant received a booklet that contained the twelve tangrams and their corresponding perspective labels. Participants had three minutes to learn the twelve perspectives, and afterwards they were given a three minute recall task without access to their study booklet. The recall task was given to ensure that they had learned the Same and Different labels. The task required them to write in an answer booklet the perspective label they had learned in a blank below each tangram. Responses were scored for correctness, and if either partner didn't achieve a perfect score, they were given additional time to learn and recall the labels until they could recall all twelve labels without error. The additional recall tasks had the tangram pictures and blanks presented in a different order.

Referential Communication Task (Continued). After the learning task was completed, participants were returned to into the experiment room and received the rest of the instructions for the RCT. Directors were instructed to describe the tangrams in the order that they appeared in their notebook, and to not skip forward or backward. Matchers were told they were to place the tangram cards on the poster board stand. They were shown where to put the first card and directed to put 6 cards on each shelf of the display board (see Figure 4). Directors were also instructed not to show any of the pictures in their booklet to matchers, and matchers were instructed not to show any of their cards to directors. Immediately before the beginning of Round 1, the tangram picture cards were randomly arranged in front of the matcher by the experimenter, and this action was repeated for each following round. While the cards were being arranged, matchers were told to take a look at them and familiarize themselves with each card and their location. Both the director and matcher were told that they were allowed to describe the pictures using the perspectives they had just learned, and that they could talk freely between each other. Each Phase II experimental session was audio taped.

2.7. Coding

Transcribing. Director's and matcher's speech during the Phase II RCT was transcribed in detail. The transcripts were segmented by trials, with notation referring to round and tangram order. For example, the first tangram that a pair described in Round 1 was designated 1–1, signifying that it was tangram number 1 in Round 1, and 2-1 was the first tangram in Round 2 and so on.

Coding Preparation. For coding purposes, what counted as a perspective was any content words (nouns, verbs, adjectives, and adverbs) spoken by either the director or matcher on their turn during a trial. Content words were recorded on a content coding sheet while leaving out "hedge" phrases like "kind of" and "sort of". To keep content provided by the director and matcher distinct, a "D:" or "M:" designation was placed next to the content words spoken by each person.

Coding. Perspectives were coded in relation to each person's prior perspective learned during the study task in Phase II. A coding scheme was used to analyze directors' and matchers' perspective setting behavior and to categorize their content words contained in their perspectives. Perspective setting behavior was coded with several dependent variables in mind (effort, balance of director's effort, accuracy, entrainment, flexibility, counterproposals, director's acceptance of counterproposals, matcher's requests for clarification). The content of their perspectives was coded with a notation scheme used to categorize perspectives based on whether they were mentioned by the director or matcher and whether they had the same, different, or no prior perspective.

2.7.1. Coding of Perspective Setting Behavior.

Entrainment. The entrainment measure captured whether a director reused a previously used expression for a particular tangram in Rounds 4 and 5 (or possibly earlier). This was a measure used to denote the earliest round in which a pair demonstrated they had converged on a particular perspective (meaning that they continued to use the same label for all following rounds). Pairs were considered to have entrained in situations where the director proposed a perspective label for a tangram that the matcher accepted with a simple acknowledgment (e.g. “ok”, “alright”, etc) and that accepted label then persisted to the final round without any errors or any radical content changes (Brennan & Clark, 1996).

Entrained perspectives were coded by two separate coders according to six sets of criteria for decreasing convergence (see Appendix A). Category 1 is an example of convergence on an exact perspective or phrase where content words used in consecutive rounds are identical in content but not necessarily order (as in the case of hybrids). For example, the hybrid phrase *the barbell dog bone* was counted as equivalent to *the dog bone barbell*. Category 2 was for expressions that were shortened, where a perspective is condensed in a following round without any new content words or phrases. Category 3 captured instances where participants re-used on a perspective used in an earlier round instead of the immediately preceding one. For my analysis, pairs of utterances coded in the first three categories were counted as instances of entrainment.

Categories 4 – 6 were where pairs were less successful in converging at the same perspective. Category 4 was where directors added or changed one or more content words between rounds. Category 5 was when partners' perspective shared some similar content words with what was mentioned in the round before but did not match in other respects, and Category 6 categorized instances where none of the content words for a perspective used between rounds were similar to any content words mentioned before. I did not count these three categories as entrainment.

Joint Effort and Distribution of Effort over the two roles. Joint effort was computed as the number of words spoken by both the director and matcher as they matched each of the tangrams in each round. I measured the proportion of the director's effort by calculating a percentage based on the dividing the number of words uttered by the director by the total number counting the percentage of words uttered by the director divided by the total amount of words spoken by both people for a specific tangram during a round.

Accuracy. Unrepaired errors were counted as the number of items placed by the matcher in the wrong location and not corrected before proceeding to the next round. Repaired errors were counted as the number of mismatched cards that were later replaced with the correct one before the end of a round.

Director's Flexibility and Matcher's Counterproposals. Flexibility was a measure designed to capture situations where the director accepted an alternative perspective provided by the matcher and they subsequently entrained on that perspective. This measure is interesting in the Different prior perspective condition. Directors demonstrated flexibility when a pair ended up with the matcher's perspective because the director typically proposed the perspective they learned first. Counterproposals are alternate perspectives given by the matcher different than the one initially proposed by the director.

Perspective content was coded for both flexibility and counterproposals using both binary (yes/no) and categorical decisions. The binary scheme was applied in two ways: 1) it was used to code the first three utterances with a particular card during a particular round. This was because when the director proposed a perspective and the matcher didn't immediately ratify it on the next turn, the rest of the trial was typically used to ground the perspective they would end up using in the future, and 2) it was also used to record any perspective setting behavior after the first three utterances to capture any additional perspective information provided by either director or matcher.

The categorical decisions were used to characterize the content words and phrases in the single and hybrid perspectives. There were thirteen perspective categories (see Appendix B), with the first four categories used for coding single perspectives and the remaining categories used for coding hybrid perspectives. A single perspective consisted of a lone content word or phrase that could be made up of previously learned or new information. A hybrid perspective was comprised of content words or phrases (most often from earlier rounds) combined into a single phrase. The category codes also indicated whether the director or matcher was the source of particular content words or parts of phrases. For instance, let's say a director studied *barbell*, and a matcher studied *dog bone* for the following tangram:

If during a particular round the director used the matcher's perspective *barbell* to refer to the tangram, then that content would be coded **M**;if the director used a hybrid perspective that consisted of their own and the matcher's prior perspectives, such as *the barbell dog bone*, then it would be coded as **D M**, as this new *hybrid* perspective contains both the director's and the matcher's prior perspectives.

2.8. Analyses

Unless stated otherwise, ANOVAS were calculated and contrasts were made among the working memory span conditions for directors' span (H_D-H_M & H_D-L_M vs. L_D-H_M & L_D-L_M, between subjects), matchers' span (H_D-H_M & L_D-H_M

vs. H_D-L_M & L_D-L_M, between subjects), and the prior perspective conditions (Same, Different, and No, within subjects). Linear trends were calculated over matching rounds (Rounds 1-5 unless stated otherwise), as well as for interactions of rounds with other variables of interest. For each statistical comparison, F_1 is the test statistic for the ANOVA aggregated by-subjects, and F_2 is by-items.

2.8.1 Analyses Notes

Data from Round 1 for a single No prior perspective tangram from a pair in the H_D-H_M span condition was removed from the analysis for the joint effort dependent measure because the number of words used in the first round for that new prior perspective item was over 6 standard deviations larger than the average number of words used by the other H_D-H_M span pairs for new items in that round. This constituted only .17 % of the data.

***No prior perspective condition* analyses.** Apart from word counts, no coding or analyses (e.g., of entrainment) were done on the tangrams from the No prior perspective condition for the Phase II dependent variables. This decision was made because there have been many referential communication studies that have analyzed perspective taking when people have had no prior perspective or knowledge of what they were trying to match, and it would be more difficult to objectively judge “whose” perspective was being entrained upon.

Gender Balance. Although the L_D-H_M condition has a disproportionate number of female-female pairs compared to the other span conditions (see Table

2), I believe that since I controlled for WM span (by assigning to hi and low groups) and the tangrams were not previously lexicalized and therefore less likely to carry any gender sensitive information. Also, the IRI scores of the directors and matchers allowed me to filter out any individuals who may also be especially resistant to taking their partner's perspective.

2.8.2 Predictions for the Prior Perspective Conditions

Predictions in this section concern the prior perspective conditions (Same, Different, and No) and the dependent measures of joint effort, proportion of director's effort, accuracy, and entrainment. Predictions about interactions of prior perspective with working memory span conditions with respect to the other perspective setting measures (counterproposals and flexibility) appear in the following section because they involve specific comparisons of the span conditions.

Joint Effort (Words). I predict that pairs will expend the most effort, and use the most turns when they have no prior perspective, followed by when they have different prior perspectives, and then the same prior perspective. I predict that people will exert the most effort when in this prior perspective condition. With no prior perspective, people will have to construct their own perspectives from scratch, exerting more cognitive effort, as they do in a typical RCT.

Proportion of Director's Effort. For all prior perspective conditions, I expect the proportion of the director's effort to decrease from rounds 1 – 5. I expect the pattern of director's effort to match that of joint effort, with director's taking the largest proportion of effort for the No prior perspective condition, followed by Different, then Same.

Accuracy. Pairs will be most accurate when they have the Same prior perspective, followed by when they have Different perspectives, and make the most errors when they have No prior perspective.

Entrainment (Different Prior Perspective Only). I expect all the pairs to entrain by Round 5. I predict an interaction with span condition, where working memory span influences the rate that people entrain on a joint perspective.

2.8.3. Primary Comparisons of Interest for the Working Memory Span Conditions

The comparisons are organized with respect to the dependent measures, with each proposed span condition comparison in parenthesis preceded by a prediction of the expected result.

Joint Effort (Words). I predict that effort for all the span conditions will decrease over the five rounds. When directors are high span (H_D-H_M or H_D-L_M), I expect them to expend less effort than the other span conditions. To answer this question, I will do a comparison of high vs. low span directors' effort (H_D-H_M & H_D-L_M vs. L-H and L_D-L_M).

Director's Proportion of Effort. The proportion of a director's effort should decrease over time. I predict that high span directors would expend a larger proportion of the effort than low span matchers (H_D-H_M & L_D-H_M vs. H_D-L_M & L_D-L_M).

Accuracy. Overall, pairs with low span matchers should make the most errors when compared to pairs with high span matchers (H_D-H_M & L_D-H_M vs. H_D-L_M & L_D-L_M), and participants in with low span people in both roles should be the least accurate (H_D-H_M & H_D-L_M, & L_D-H_M vs. L_D-L_M) because high span matchers would be able to accommodate their responsibility of minimizing errors.

Entrainment (for the Different prior perspective condition). Pairs with high span matchers should entrain earlier than pairs with low span matchers (H_D-H_M & L_D-H_M vs. H_D-L_M & L_D-L_M) because high span matchers would be better able to accommodate the directors' perspective along with their own, and therefore make it possible for pairs to entrain in an earlier round.

Flexibility (for the Different prior perspective condition). Pairs with high span directors should demonstrate more flexibility that pairs with low span directors (H_D-H_M & H_D-L_M vs. L_D-H_M & L_D-L_M). The greatest flexibility should be when high span directors are paired with low span matchers (H_D-L_M vs. all other span conditions), high span directors with low span matchers should demonstrate the most flexibility out of all the directors in the experiment, because they would have the most opportunities to adjust perspectives, due to the potential difficulty low span matchers may have adapting to their partner's perspectives (high span directors with high span partners would have less of a need to adjust). Low span directors should show the fewest examples of flexibility.

Counterproposals (for the Different prior perspective condition). Low span matchers will give more counterproposals than high span matchers (H_D-L_M & L_D-L_M vs. H_D-H_M & L_D-H_M), because they will be less able to accommodate the director's perspective, and be more inclined to volunteer their own.

Chapter 3. Results

3.1. Joint Effort and Accuracy

A pair's total effort was estimated by the combined number of words used by the director and matcher for each of the five rounds (as in Bortfeld & Brennan, 1997); directors' relative proportion of effort was estimated as the number of the director's words divided by the pair's total words. Table 3 shows the mean number of words spoken for both directors and matchers combined. Errors were determined by counting the time that matchers placed a picture card in the wrong position on the target area and did not change it before proceeding to the next round.

Accuracy. Overall, Figure 5 shows that errors decreased over time, linear trend, $F_1(1,28) = 13.10$, $p < .01$, $F_2(1,15) = 23.67$, $p < .001$. As predicted, accuracy in the conversational task was generally high, yet was influenced by prior perspective condition, $F_1(2,56) = 6.63$, $p < .05$, $F_2(2,15) = 3.82$, $p < .05$, and pairs with low span matchers (H_D-L_M & L_D-L_M) made more errors overall than pairs with high span matchers (H_D-H_M & L_D-H_M), F_1 $(1,30) = 5.58$, $p = .03$, F_2 $(1,15) = 3.35$, $p = .09$.

Although pairs made the most errors when they had No prior perspective compared to when they had the Same or Different one, $F_1(1,30) = 5.96$, $p = .021$, $F_2(1,15) = 5.56$, $p = .03$, their accuracy was also influenced by an interaction between their prior perspective and the matchers' working memory span, F_1

(2,60) = 2.60, $p = .08$, F_2 (2,15) = 3.82, $p < .05$. In particular, low span matchers were also slower to reach peak efficiency (minimal to no errors while using few words) over 5 rounds when they had a Different prior perspective (vs. Same), linear trend, F_1 (1,28) = 5.50, $p < .03$, F_2 (1,15) = 3.38, $p = .06$.

Joint Effort (Words). Indeed, the prior perspectives on an item affected the effort pairs expended together to successfully refer to it, $F_1(2,56) = 40.86$, $p < .001$, $F_2(2,15) = 5.87$, $p = .01$. That is, they expended the least effort when they had the Same prior perspective compared to Different prior perspectives, $F_1(1,28) = 28.28$, $p < .001$, $F_2(1, 15) = 4.98$, $p = .09$, and the most effort when they had No prior perspective as compared to Different perspectives, $F_1(1,28) = 17.84$, $p < .001$, F_2 (2,15) = 4.98, $p < .05$ (see Figure 6). Pairs in all four span conditions became more efficient over time, linear trend, F_1 (1,28) = 68.89, $p < .001$, $F_2(1,15) = 66.42$, $p < .001$.

There was no main effect of span condition (H_D-H_M, H_D-L_M, L_D-H_M, L_D-L_M) for total word counts by subjects, $F_1(3,28) = 1.53$, $p = .228$, n.s., $F_2(3,45) = 7.611$, $p <.001$. However, pairs' total word counts were influenced by an interaction between their prior perspective and the matcher's working memory span, F_1 (2,60) = 3.33, $p < .05$, $F_2(2,15) = 3.71$, $p < .05$, especially when they had No prior perspective compared to when they had any prior information (Same or Different), $F_1(1,30) = 4.35$, $p = .047$, $F_2(2,15) = 5.66$, $p = .02$. Pairs that consisted of low span directors and high span matchers expended less joint effort (marginally by-subjects and reliably by-items) than pairs where both people were low span (L_D-H_M vs. L_D-L_M), $F_1(1,28) = 4.37$, $p < .096$, $F_2(1,15) = 21.76$, $p < .001$, and low span pairs (L_D-L_M) expended the most effort overall (L_D-L_M vs. the other three span conditions combined), $F_1(1,28) = 3.57$, $p < .07$, $F_2(1,15) = 10.50$, $p < .005$.

Distribution of Effort over the Two Roles (Proportion of Directors' Effort). I used this measure to see whether two partners distributed their respective effort differently in the face of different task difficulty (Same vs. Different vs. No), over time (5 rounds), or depending on either partner's memory span. The proportion of words produced by the director is one way of capturing relative effort in different task roles. Overall, the proportion of words produced by directors was influenced by prior perspective, $F_1 (2,56) = 17.36$, $p < .001$, $F_2(2,15) = 5.46$, $p = .02$. That is, directors expended higher proportions of effort when pairs had the Same prior perspective compared to when they had Different ones, $F_1(1,28) = 3.59$, $p < .07$, $F_2(1,15) = 5.46$, $p <.006$, and a still higher proportion of effort with No prior perspective (vs. Different), $F_1(1,28) = 24.66$, $p < .01$, $F_2(1,15) = 5.48$, $p = .04$. This may have happened because in the different perspective condition, matchers expended a higher proportion of effort than in other two perspective conditions because they often needed to convey or propose their learned perspectives to the directors (and this was not the case in the other two conditions, where directors did most of the perspective proposing. Also, that directors should take on more of the effort that pairs expend jointly when their matchers had no prior perspective is in agreement with my predictions, as they would have to work harder to describe and propose a joint perspective when neither partner has anything to start with (see Figure 7).

The directors' proportion of effort decreased over rounds, linear trend, $F_1(1,28) = 5.89$, $p < .05$, $F_2 (1,15) = 3.66$, $p = .08$. This interacted with prior perspective, $F_1(1,28) = 2.35$, $p < .05$, $F_2(1,15) = 3.02$, $p < .01$. When they had the Same prior perspective, directors decreased their proportion of effort, linear trend, $F_1(1,28) = 11.09$, $p < .002$, $F_2(1,15) = 13.30$, $p < .01$. However, when they had Different prior perspectives, the distribution of effort changed in Round 2 to reflect an increase for the director and then decreased over the final 3 rounds, quadratic trend, $F_1 (1,28) = 20.90$, $p < .001$, $F_2(1,15) = 8.52$, $p =.003$, and the pattern for distribution of effort was similar when pairs had No prior perspective, quadratic trend, $F_1 (1,28) = 3.06$, $p = .09$, $F_2 (1,15) = 13.96$, $p < .01$.

A closer look revealed that directors' and matchers' working memory span interacted with their prior perspective condition, and this influenced how they distributed their effort over time. The linear trend over rounds of the director's proportion of effort for the H_D-H_M pair when they had No prior perspective and the L_D-L_M pairs when they had a Different prior perspective illustrated this interaction. Directors in H_D-H_M pairs had two *increases* in their proportions of effort when they had No prior perspective with their partner, from rounds 1-2 and

4-5, and directors in L_D-L_M pairs, when they had Different prior perspectives, had two *decreases* between rounds 1-2, and rounds 4-5, cubic trend, $F_1(1,28) = 3.06$, $p =.09$, $F_2(1,15) = 11.62$, $p = .02$.

3.2. Directors' and Matchers' Perspective Setting Behavior

Results for the entrainment measures are reported for Different prior perspectives only, because with Same prior perspectives, pairs entrained at uniformly high rates. Accordingly, the dependent measures of flexibility, matcher's counterproposals, director's acceptances, and subsequent mentions of the matcher's learned perspective are of interest for only the Different perspective condition.

Entrainment. Pairs nearly always entrained and committed to a joint perspective for the 6 tangrams steadily increasing over rounds, $F_1 (1,28) = 163.99$, $p <.001$, $F_2 (1,10) = 126.45$, $< .001$; this was true for all span conditions (Figure 8a). But span did appear to shape the patterns with which entrainment increased over rounds, as evident in the superimposed lines in Figure 8b. As predicted, working memory span had an effect on how pairs entrained. When entrainment is contrasted round by round for the low span matchers, Round 2 is where the H_D-L_M pairs' rates of entrainment begin to exceed those of the L_D-L_M pairs, $F_1 (1,28) = 4.17$, $p < .05$, $F_2 (1,10) = 4.75$, $p = .07$, and begins to approach that of the high span matchers. But the L_D-L_M pairs' rate of entrainment doesn't approach the level of the others until Round 3, the last round where their performance still lags behind the other span pairs, F_1 (1,28) = 4.27, p =.07, F_2 (1,10)= 5.29, $p < .05$ (see Table 4 and Figure 8b).

Flexibility. Figure 9 shows the instances of flexibility in perspective taking demonstrated in Round 5 by the directors from each span condition. I estimated flexibility by evaluating the perspective pairs had entrained on by Round 5 with respect to two criteria: 1) it was a perspective that they had used in Round 4, and 2) the perspective was the matcher's originally-learned perspective. I anticipated that high span directors would demonstrate more flexibility than low span directors, and although this difference was not significant, the means were in the right direction (H_D-H_M mean = 1.8 per round, H – L mean = 2.7 per round, L_D-H_M mean = 1 per round, and L_D-L_M mean = 1.7 per round), F_1 (1,28) = 1.838, p = .186, n.s., F_2 (1,10) = 1.90, $p < .197$, n.s. I also predicted that the H_D-L_M pair would show the most flexibility compared with all the other span pairs, as low-span matchers should have the most trouble abandoning their previously learned perspectives in order to accept the directors' proposals, and high-span directors should be more likely to have the resources to abandon their own learned perspectives. While this difference was not reliable, the means also went in the predicted direction (H_D-L_M vs. H_D-H_M, H_D-L_M, & L_D-L_M, means of 2.7 per round vs. 1.5, respectively), F_1 (1,30) = 2.12, p = .156, n.s., F_2 (1,10) = 2.98, p= .115, n.s.

Matchers' Counterproposals. Figure 10 shows that although matchers generally offered fewer counterproposals over time, linear trend, F_1 (1,28) = 17.48, $p < .001$, F_2 (1,10) = 27.03, $p < .001$, high span matchers offered low span directors more counterproposals than matchers in all the other span conditions (L_D-H_M vs. all the other span conditions combined), F_1 (1,30) = 2.97, $p < .096$, F_2 (1,10) = 34.508, $p < .001$.

Directors' Acceptances of Matchers' Counterproposals. Directors' working memory span influenced their likelihood of accepting matchers' counterproposals, with high span directors accepting the most per round (H_D-H_M & H_D-L_M > L_D-H_M & L_D-L_M), F_1 (4,112) = 2.31, $p < .04$, F_2 (4,40) = 2.492, $p < .06$, even though low span directors with high span matchers had more opportunities overall to accept matchers' counterproposals (see Figure 11). When these acceptances were normalized for the number of counterproposals that matchers made, high span directors also accepted the highest proportion per round (H_D-H_M & H_D-L_M > L_D-H_M & L_D-L_M), F_1(4,112)=2.31, $p <.06$, F_2 (4,20) =3.13, $p <.04$.

Matcher's Requests for Clarification after the Director's 1st Description. Not surprisingly, matchers made fewer requests for clarification over time (Figure 12), F_1 (1,28) = 96.44, $p < .001$, F_2 (1,10) = 48.93, $p < .001$. Pairs in all span conditions made fewer requests for clarification when they had the Same prior perspective than Different prior perspectives, F_1 (1,28) = 39.51, $p < .001$, F_2 (1,10) = 7.47, $p < .05$.

Mirroring their pattern of counterproposals, high span matchers increased their requests for clarification from their low span directors over the last two rounds, cubic trend, F_1 (1,28) = 3.93, $p < .06$, F_2 (1,10) = 6.89, $p < .05$.

Directors' Elaborations on their Initial Proposal. I counted as elaborations instances where the director said more than their learned perspective when describing a tangram on their first turn. Directors elaborated less over time, linear trend, F_1 (1,28) = 134.16, $p < .001$, F_2 (1,10) = 34.48, $p < .001$. When they had Different prior perspectives, directors elaborated more often than when they had the Same one, F_1 (1,28) = 31.76, $p < .001$, F_2 (1,10) = 6.650, $p < .03$ How likely directors were to elaborate depended marginally on both their own and their matchers' working memory spans, F_1 (1,28) = 2.88, $p = .10$, F_2 (1,10) = 3.970, $p < .02$. That is, directors elaborated most when both the director and matcher were low span and had Different prior perspectives, F_1 (3,28) =3.513, $p = .06$, F_2 (1,10) = 5.45, $p < .05$ (see Figure 13).

Director's use of Hybrid Perspectives on their 1st Turn (Rounds 2-5). Proposing a hybrid perspective (e.g. "barbell dog bone") is another way in which directors could take account of matchers' needs, if they're aware that matchers have studied Different perspectives. Overall, directors in all span conditions decreased the number of hybrids they used over time, F_1(1,28) = 14.48, $p < .001$,

F_2 (1,10) = 11.413, $p < .008$ (see Figure 14). As expected, directors also used more hybrids when they had Different prior perspectives compared to when they had the Same one, reliable by-subjects but not by-items, F_1 (1,28) = 10.12, p = .004, F_2 (1,10) = 2.30, p = .160, n.s. When directors had low span matchers, they increased the number of hybrids they used from Round 1-2, F_1(1,30) = 3.38, $p < .08$, F_2 (1,10) = 5.44, $p < .05$.

Mention of Matchers' Perspectives (Rounds 2-5). Overall, the matchers' learned perspectives were mentioned less over time, linear trend,F_1 (1,28) = 24.55, $p < .001$, F_2 (1,10) = 5.09, $p < .05$, and the linear trend of whether matchers mentioned their own perspectives was marginally influenced by their own and the directors' working memory span, F_1 (1,30) = 2.16, $p < .08$, F_2 (1,10) = 4.28, $p < .07$. The linear trend of the L_D-H_M pair shows that high span matchers mentioned their own perspective more often over the last two rounds when they had low span directors, quadratic trend, F_1(1,28) = 3.84, $p < .06$, F_2 (1,10) = 5.10, $p < .05$ (see Figure 15a, Figure 15b).

Chapter 4. Discussion

The goal of this dissertation project was to initiate a research program investigating how an individual cognitive difference like working memory span might shape referential communication. Referential communication is a primary function of language use and an omnipresent one; it is not difficult to find situations where people are collaborating and trying to coordinate their mutual knowledge about objects in the world. When people refer to an object, they often collaborate to set a joint perspective using terms that they can use to re-refer to that object in the future (Schober & Clark, 1989; Wilkes-Gibbs & Clark, 1992). While setting a joint perspective, people have the task of dynamically monitoring and managing multiple sources of information (e.g. the situation, the partner, the goals of the conversation) as they plan and comprehend what is said. People have different cognitive capacities for accommodating and managing information moment by moment in working memory, and their perspective setting behavior should be affected by how they perform cognitive subtasks associated with referential communication (evaluating differences among objects, mapping referring expressions onto objects, repairing misunderstandings) while in their respective roles (director, matcher).

Directors' working memory span and perspective setting. Directors with high working memory spans had cognitive resources that allowed them to be more efficient in more demanding situations than directors with low working memory spans. While setting a joint perspective, directors spent most of their time dealing with any misunderstandings the matcher may have had with respect to their initial proposed perspective, and mapped the matcher's feedback onto what they just proposed to craft a referring expression that the matcher would accept. These conversational subtasks became more or less difficult, depending on the amount of prior perspective information that they shared with their matcher. When they had Different perspectives from matchers, at least directors had some information to start with, and it was easier to modify a prior perspective on the fly than to start from scratch with No prior perspective. Directors expended more effort (and a greater proportion of the total pairwise effort) when they had to create a new perspective on the fly and then propose it to matchers and get it accepted.

High span directors seemed more able to manage the resource demands of the conversational context and perform adjustments to their matchers' behavior. Low span directors were able adjust to matchers too, but with adjustments that can be thought of as less cognitively demanding. Some types of adjustments in particular could be categorized as being relatively easy on working memory (e.g. hybrid use on the 1st turn and elaborating on the 1st proposal in subsequent rounds) as opposed to more challenging (e.g. taking the matcher's perspective and accepting the matcher's counterproposals).

With respect to the easier types of potential adjustments directors could make, elaborating on their initial proposal in later rounds is easy because directors have only to provide additional information on their first turn in addition to their learned perspective (e.g. “the barbell with the diamond shapes on the side”). It is striking that the most cases of elaborations occurred with low span directors interacting with low span matchers. Another relatively easy adjustment for directors to make, using a hybrid perspective, requires only proposing their own perspective along with the matchers’ perspective, if the matchers revealed it previously (e.g. “barbell dog bone”). The fact that low span and high span directors demonstrated similar patterns of hybrid usage over time and across perspective conditions for low span matchers is evidence that this type of potential adjustment might not be that difficult for directors to make.

One of the more challenging director adjustments, adopting the matcher’s perspective, requires directors 1) to overcome the natural tendency of the RCT that favors the director’s perspective being the joint one, because it is almost always proposed first (as the director is the one who knows the target arrangement) and 2) to accept a counterproposal, map it onto the one they already have, and then abandon their own learned perspective. The results of this experiment suggest that although there were not many examples of directors adopting the matcher’s perspective, they still attempted to do so, and this was done almost exclusively by high span directors.

With respect to the easier types of director adjustments, low span directors demonstrated similar patterns of hybrid usage as did high span directors for low span matchers across prior perspective conditions and over time. However, high span directors accepted more counterproposals from their matchers; low span directors might have had more trouble with that more challenging adjustment. Not only is accepting a counterproposal and giving up a prior perspective likely to be more difficult than crafting a hybrid, but creating a hybrid perspective with might ease the cognitive burden for both partners by combining the cues from both perspectives (e.g. "the barbell dog bone.").

Matcher's working memory span and perspective setting. Possessing a high working memory span might allow matchers to use more efficient searching and comparison strategies when they are scanning the item array. Since a person with a high memory span may be able to maintain many items, categories, or descriptions in working memory, they might be able to increase the efficiency with which they encode the different qualities of the objects and the director's referential terms, enabling them to map the directors referring expression onto objects more effectively. Matchers have to map the director's referring expression onto their item array and evaluate the differences among objects to reduce the probability of selecting the wrong item. This job was made more or less difficult based on the amount of prior perspective information that they shared with their matcher. When they had the Same prior perspective, it was easy. When they had

Different prior perspectives to start with, it was more difficult, but at least directors and matchers had some form of a representation to work with. With No prior perspective, the cognitive tasks for both directors and matchers were most difficult because they had to create new perspectives on the fly, and then test these in communication with one another.

Pairs with high span matchers may have been able to perform referential subtasks more effectively, and benefit from the increased joint effort early on that usually facilitated relatively early entrainment. Relatively earlier entrainment was associated with fewer errors, and low span matchers may have had an effort / accuracy tradeoff suggesting that they were less able to judge how much effort they needed to put in to succeed. Pairs with low span matchers made more errors and did not correct them later, even in later rounds when things should have been easier.

Matchers during conversation can also deliver evidence of their understanding that can make conversation more efficient, and the potential techniques for doing this can be categorized as being less taxing to perform (requesting clarification) or more taxing (offering counterproposals). When they make a request for clarification, matchers can give many simple indications that they do not understand, such as a simple “huh?” or a “what?” to indicate their level of understanding.

Crafting a counterproposal could have been difficult for low span matchers to do; this is suggested by the finding that high span directors ended up accepting more counterproposals than low span directors. It could be the case that crafting a counterproposal, even when they have their own perspective, could represent a concurrent cognitive task that increases cognitive load while setting a joint perspective.

Directors' and matchers' working memory span together. With respect to the mention of matchers' perspectives over time, the pattern seems to reflect a combination of the directors' effort to accommodate counterproposals in interaction with the matcher's ability to offer them. High span directors accepted more counterproposals overall, and this may have led high span matchers to offer more, especially if directors accommodated their perspectives early in the task (when they were expending more effort and perspective-setting was more difficult). With low span matchers, the directors' ability to accommodate their matchers' perspective seemed to drive them to mention the matcher's perspective more often even when the matcher was not offering it as a counterproposal. For low span directors, it appears that when their partners mentioned their own perspective (as in the L_D-H_M pair), they made such an adjustment, but the matcher had to continue to provide counterproposals over time to sustain their mention. Although it seems that just the act of providing counterproposals can prompt a low span director to mention the matcher's perspective over time, these directors still had trouble accepting the matchers' perspectives (and in doing so, demonstrating flexibility). With a low span matcher (L_D-L_M), low span directors do not have partners who can offer counterproposals (because they don't have the working memory span to accommodate it) at a rate that might prompt directors to use the matcher's perspective more over time.

The interactions of working memory span (high/low) with task roles (director/matcher) on various measures provide insights about how pairs adjusted their balance of effort according to task demands. Although the matching task becomes easier over rounds, the cubic trend of the H_D-H_M and the L_D-L_M pairs reveal that there could be different task demands in the early or late rounds. In the early rounds, the matcher must work hard to map expressions onto items in the array, and in H_D-H_M pairs, director's seem to exhibit an increase in their relative effort by saying more words that may accommodate the matcher's needs. In the late rounds, directors have the challenge of managing a series of committed perspectives that began with different degrees of lexicalization, and the H_D-H_M pairs may have demonstrated adjustments in their balance of effort in response to the director's round to round needs in that respect, while the L_D-L_M pairs appeared to show changes in the opposite direction. Although the effort and cognitive workload in the late stages in the conversational task are much lower, it appeared that high span pairs made more round-to-round adjustments over time to their distribution of effort with the No prior perspective items than did the low span pairs. Although low span pairs appeared to make large adjustments to the director's proportion of effort in response to *both* their informational needs, such as when they both might have some prior information (e.g. when they have the Same or Different prior perspective), they seem to have a more difficult time making adjustments to their balance of effort in response to either partner's needs separately, even in the late rounds when it is easier to do so.

I also predicted that when both members of a pair had a high span (H_D-H_M) in Phase II, they would be more efficient (e.g. fewer errors and words over time). The idea was that having high span people in both roles would make their referential communication more efficient; that is, their greater working memory capacity would make them expend less effort to reach a joint perspective. However, this did not happen. Pairs with high span matchers (H_D-H_M, L_D-H_M) were more efficient than pairs with low span matchers (H_D-L_M, L_D-L_M). It could have been the case that pairing two high span people did not provide an additive boost to efficiency because working memory span may be more important in the matcher's role: Matchers have the primary job of monitoring for and initiating the repair of errors. A high span in the director role did not appear to enhance how quickly pairs converged on a joint perspective.

Conclusions and Implications. Individual differences in working memory span do affect perspective taking. As predicted, a person's working memory span influenced their capacity to contribute efficiently, as well as the types of adjustments they made while perspective setting with another person whose perspective contrasted with their own.

The impact of working memory span was role specific with regard to audience design. Working memory span affected how well directors and matchers retrieved memories about what salient information is shared with their partners as well as how to use that information in conversation) (see Horton & Gerrig, 2005a, for discussion of these component processes). Directors engaged in audience design exerted most of their effort toward creating an effective description, and matchers put most of their effort into monitoring when they thought a match was made and determining when they could move on to the next item.

Although span did not affect how rapidly directors established joint perspectives with their partners, it did affect the strategies they used to do so. These strategies can be explained by differences in resource demand on working memory. While high span directors were more likely to make costly adjustments like abandoning their own perspective and adopting their partner's, low span directors also took their partners into account. They did it in less costly (more incremental) ways, by using hybrid perspectives or elaborating on their initial description.

As for matchers, both high and low span matchers also appeared to use audience design strategies that had different costs to their working memory. High span matchers more likely to offer feedback that contained an alternative perspective while low span matchers were less likely to do so. Note that participants were unaware of their partners' working memory span classification, and since they did not switch director-matcher roles during the experiment, they were unaware of how difficult their partner's role was. This suggests that adjustments were more likely to have been made on the basis of their partner's feedback rather than based on a stereotyped model of their partner's ability.

The interactive nature of dialogue provides interlocutors with multiple opportunities for expression and understanding, making it robust and forgiving, and also sets the stage for efficient resource allocation. People can adjust their effort based on what they perceive is needed moment by moment, as the demands on their cognitive resources by the conversation change over time. The present experiment showed expending effort early on ensured that by the fifth round, partners converged on a shared perspective. The results imply that adults with a wide variety of working memory spans can make some kind of adjustments while setting a perspective. This may be good news for the elderly and others who have a working memory span limitation; future work should focus on such individuals. Another avenue for future work is whether perspective-taking strategies can be learned via specific training or experience. Working memory research has demonstrated that span does not determine how well people can adopt task strategies (Schunn & Reeder, 2001). If people with different working memory spans were required to switch roles during the experiment, people with a range of spans may learn to achieve and use joint perspectives more rapidly and efficiently with their partners.

Individual differences in working memory span may not lead to simple or additive advantages, in the sense that “more is better”, but they do lead to measurable differences in how people use language during conversation.

Bibliography

Baddeley, A.D. (2002). Is working memory still working? *European Psychologist, 7*, 85–97.

Baddeley, A. D. (1994). Working memory: The interface between memory and cognition. In D. L. Schacter & E. Tulving (Eds.), *Memory Systems*. Cambridge, MA: MIT Press.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G.H. Bower (Ed.), *The Psychology of Learning and Motivation, Vol. 8*, New York: Academic Press.

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

Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology 18,* 355–387.

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.

Braver, T.S., Cohen, J.D., Nystrom, L.E., Jonides, J., Smith, E.E., & Noll, C.D. (1997). A parametric study of prefrontal cortex involvement in human working memory. *NeuroImage, 5,* 49-62.

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., & Lockridge, C. B. (2005). *How visual copresence and joint attention shape speech planning*. Unpublished manuscript.

Brennan, S. E., & Lockridge, C. B. (2006). Computer-mediated communication: A cognitive science approach. In K. Brown (Ed.), ELL2, *Encyclopedia of Language and Linguistics, 2nd Edition* (pp. 775-780). Oxford, UK: Elsevier Ltd.

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

Cahn, J. E., & Brennan, S. E. (1999). A psychological model of grounding and repair in dialog. Proceedings, *AAAL fall symposium on psychological models of communication in collaborative systems*, 25-33. North Falmouth, MA: American Association for Artificial Intelligence.

Caplan, D. (1992). *Neurolinguistics and linguistic aphasiology: An introduction*. Cambridge: Cambridge Press.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology, 4*, 55-81.

Clark, H. H., and Krych, M. A. 2004. Speaking while monitoring addressees for understanding. *Journal of Memory and Language, 50*, 62-8 1.

Clark, H. H., & Marshall, C. R. (1981). Definite reference and mutual knowledge. In A. K. Joshi, B. L. Webber, & I. A. Sag (Eds.), *Elements of Discourse Understanding* (pp. 10-63). Cambridge: Cambridge University Press.

Clark, H. H., & Schaefer, E. F. (1987). Concealing one's meaning from overhearers. *Journal of Memory and Language , 26*, 209-225.

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

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning & Verbal Behavior, 19,* 450-466.

Davis, M. H. (1983). Measuring individual differences in empathy: evidence for a multidimensional approach. *Journal of Personality and Social Psychology, 44, 1*, 113-126.

Delis, D.C., Freeland, J., Kramer, J.H., Kaplan, E. & Ober, B.A. (1987). *The California Verbal Learning Test*. New York: The Psychological Corporation.

Doherty-Sneddon, G., Anderson, A., O'Malley, C., Langton, S., Garrod, S., & Bruce, V. (1997). Face-to-face and video-mediated communication: A comparison of dialogue structure and task performance. *Journal of Experimental Psychology :Applied, 3, 2,* 105-125.

Elffers, J. & Schuyt, Ml. (1997). *Tangram: The ancient Chinese puzzle*. New York: Stewart, Tabori and Chang.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A.(1999). Working memory, short-term memory, and general fluid intelligence: A latent variable approach. *Journal of Experimental Psychology: General, 128*, *3*, 309-331.

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

Garrod, S. & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, *27*, 181-218.

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

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

Gobet, F., & Simon, H. A. (1996). Recall of random and distorted chess positions: Implications for the theory of expertise. *Memory & Cognition*, *24*, 493-503.

Hannah, J. (2001). *The effects of linguistic form, common ground, and perspective on domains of referential interpretation*. Unpublished Doctoral Dissertation. University of Rochester.

Hanna, J.E. & Tanenhaus, M.K. (2003). Pragmatic effects on reference resolution in a collaborative task: Evidence from eye movements. *Cognitive Science*, *28*, 105-115.

Horton, W. S., & Gerrig, R. J. (2002). Speakers' 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., & 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., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition, 59*, 91-117.

Issacs, E. A., & Clark, H. H. (1987). References in conversation between experts and novices. *Journal of Experimental Psychology: General, 116(1)*, 26-37.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99, 1,* 122-149.

Kane, M. J., & Engle, R. W. (2003). Working memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to stroop interference. Journal *of Experimental Psychology: General*, *132*, 47-70.

Kane, M. J., Conway, .A.R.A., & Engle, R.W. (1999). What do working memory tests really measure?: A comment on Caplan and Waters. *Behavioral and Brain Sciences, 22*, 101-102.

Kane, M.J., Bleckley, K.M., Conway, A.R.A., & Engle, R.W. (2001). A controlled-attention view of working memory capacity. *Journal of Experiment Psychology: General*, *130,* 169-283.

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

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

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

Krauss, R., & Glucksberg, S. (1969). The development of communication: Competence as a function of age. *Child Development, 40*, 255-266.

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

Kraut, R. E., Gergle, D., & Fussell, S. R. (2002). The use of visual information in shared visual spaces: Informing the development of virtual co-presence. *Proceedings of CSCW 2002*. NY: ACM Press.

La Pointe, L. B., & Engle, R. W. (1990). Simple and complex word spans as measures of working memory capacity. *Journal of Experimental Psychology: Learning, Memory and Cognition, 16*, 1118-1133.

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

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

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

Miller, G. A. (1956). The magical number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63,* 81-97.

Miyake, A., Carpenter, P. A., & Just, M. A. (1994). A capacity approach to syntactic comprehension disorder: Making normal adults perform like aphasic patients. *Cognitive Neuropsychology, 11*, 671–717.

Norman, M.A., Evans, J.D., Miller, W.S., & Heaton, R.K. (2000). Demographically corrected norms for the California Verbal Learning Test. *Journal of Clinical and Experimental Neuropsychology, 22, 1*, 80-94

Palladino, P., Cornoldi, C., De Beni, R., & Pazzaglia, F. (2001). Working memory and updating processes in reading comprehension. *Memory & Cognition, 29, 2*, 344-354.

Pearlmutter, N. J., & MacDonald, M. C. (1995). Individual differences and probabilistic constraints in syntactic ambiguity resolution. *Journal of Memory and Language, 34*, 521-542.

Pickering, M.J. & Garrod, S. (2004) Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences. 27,* 169-225.

Ragland, J. D., Turetsky, B.I., Gur, R.C., Gunning-Dixon, F., Turner, T., Schroeder, L., Chan, R., & Gur, R.E. (2002). Working memory for complex figures: An fMRI comparison of letter and fractal n-back tasks. *Neuropsychology, 16, 3*, 370-379.

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

Schober, M. F. (1995). Speakers, addressees, and frames of reference: Whose effort is minimized in conversations about locations? *Discourse Processes, 20*, 219-247.

Schober, M. F. (1998b). *How partners with high and low spatial ability choose perspectives in conversation*. Paper presented at the 39th annual Meeting of Psychonomic Society, Dallas, TX.

Schober, M. F., & Brennan, S. E. (2003). Processes of interactive spoken discourse: The role of the partner. In A. C. Graesser, M. A. Gernsbacher, & S. R. Goldman (Eds.), *Handbook of discourse processes* (pp. 123-164). Hillsdale, NJ: Lawrence Erlbaum.

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

Schunn, C. D., Lovett, M. C., & Reder, L. M. (2001). Awareness and working memory in strategy adaptivity. *Memory & Cognition, 29, 2*, 254-266.

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General, 125,* 4-27.

Staver, J. R., & Jacks, T. (1988). The influence of cognitive reasoning level, cognitive restructuring ability, disembedding ability, working memory capacity, and prior knowledge on students' performance on balancing equations by inspection. *Journal of Research in Science Teaching, 25,* 763-775.

Stellmann, P., & Brennan, S. E. (1993). *Flexible perspective-setting in conversation*. Paper presented at the 34th annual meeting of the Psychonomic Society, Washington, DC.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. E. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268,* 632-634.

Trueswell, J. C., Sekerina, I., Hill, N., & Logrip, M. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition*, *73*, 89-134.

Tulholski, S.W., Engle, R.W., & Baylis, G.C. (2001). A controlled-attention view of working memory capacity. *Journal of Experiment Psychology: General*, *130*, 169-183.

Wilkes-Gibbs, D. (1986). *Collaborative processes of language use in conversation.* Unpublished doctoral dissertation, Stanford University.

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

Wilkes-Gibbs, D., & Kim, P. H. (1991). *Discourse influences on memory for visual forms*. Paper presented at the 1991 meeting of the Psychonomic Society, San Francisco, CA.

Figure 1. Example of item on Tangram Norming Task.

"Barbell" or "Dog bone"

1	2	3	4	5
can't see It at all	can see with difficulty	can see somewhat	can see reasonably well	can see easily

Figure 2. Distribution of Operation-Span Scores for all participants in Phase II

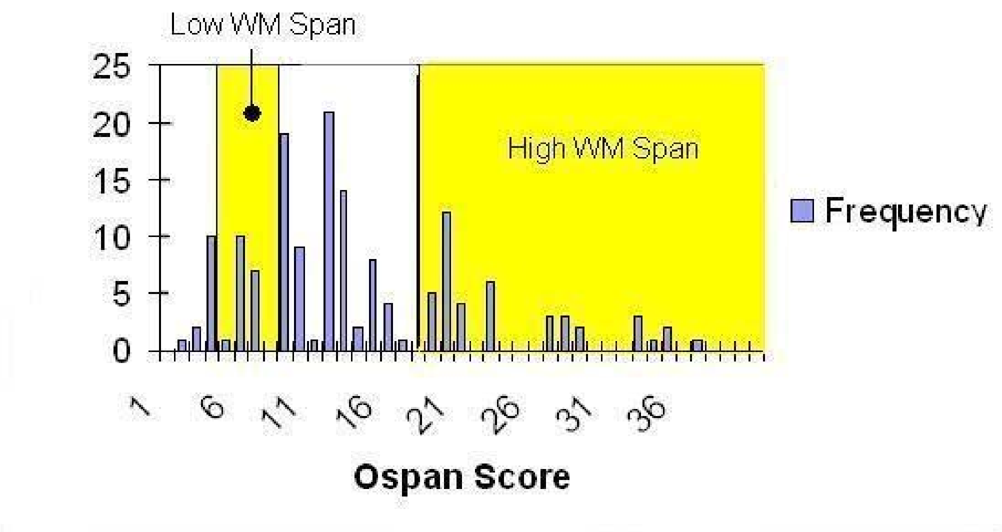


Figure 3. Example of Tangram Picture for Different Prior Perspective Condition.

Figure 4. Overhead Diagram of Phase 2.

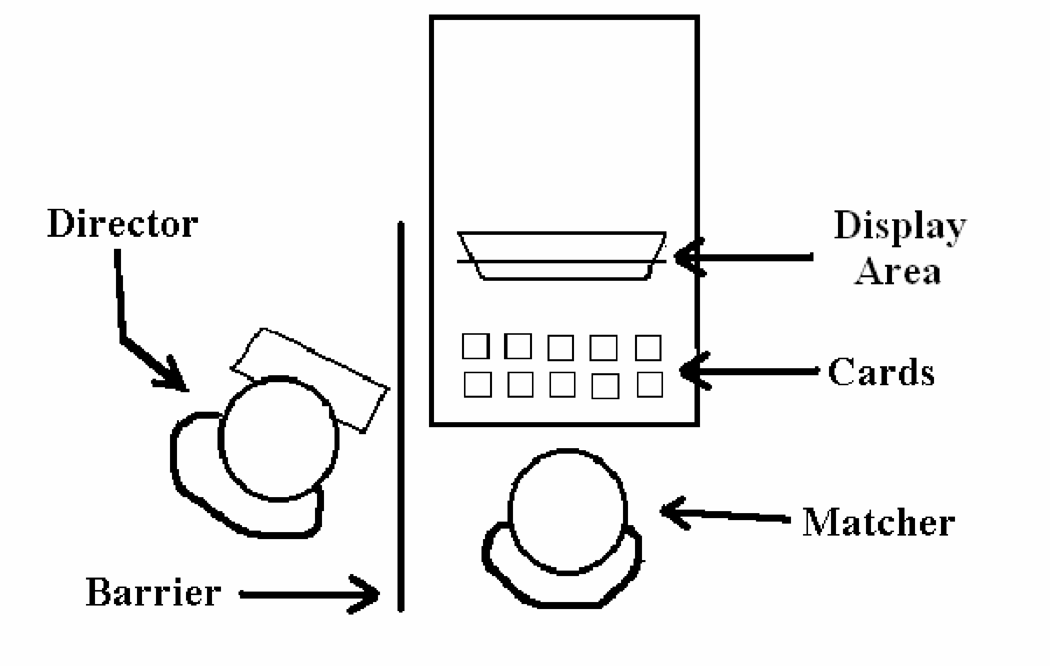


Figure 5. Errors by round, for each prior perspective and Span condition.

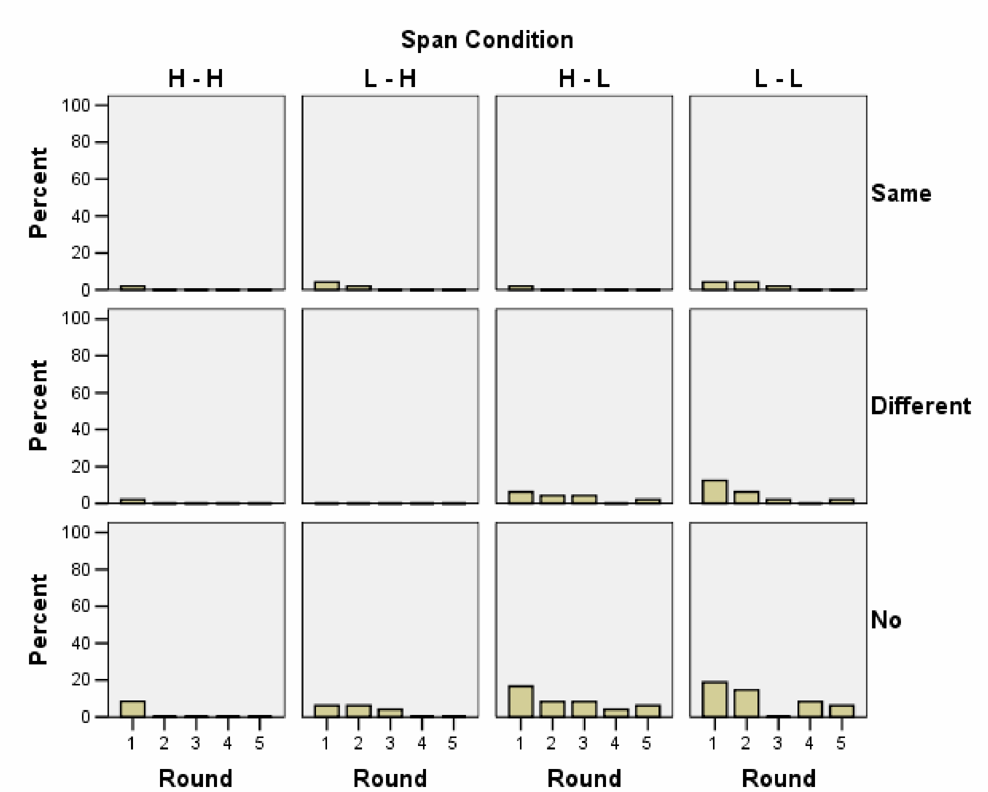


Figure 6. Mean word counts over 5 rounds, for each prior perspective condition and span condition.

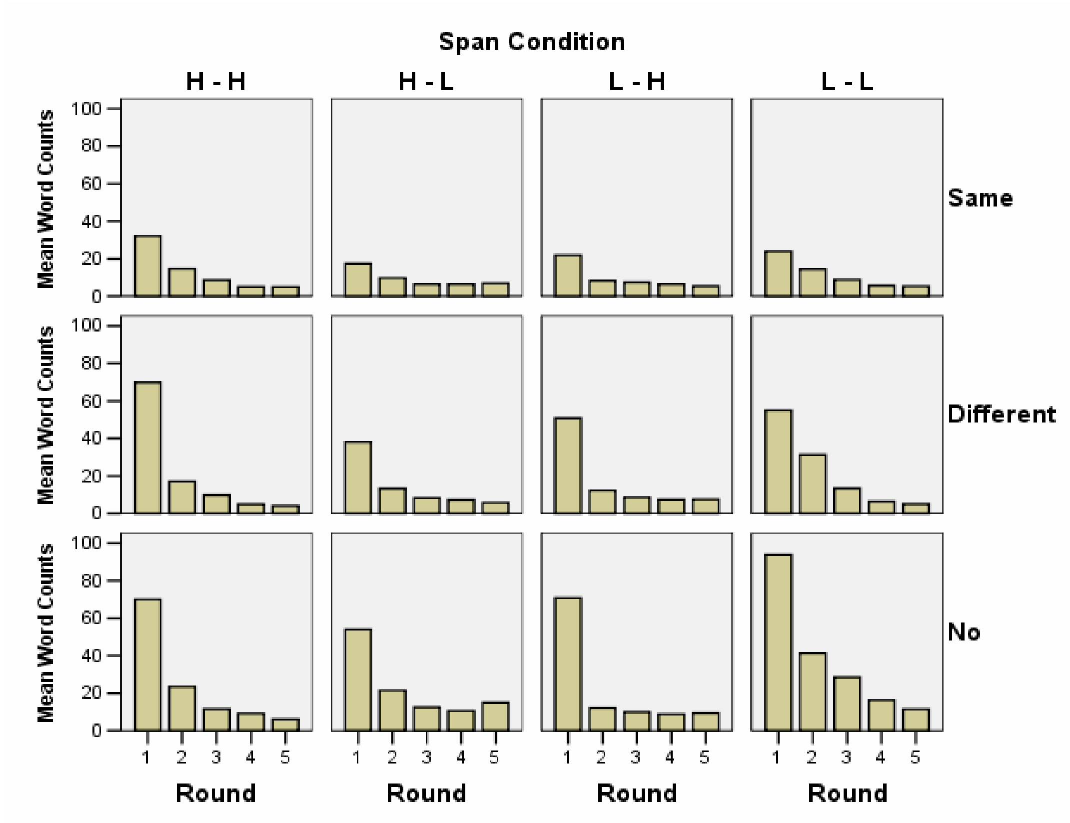


Figure 7. Director's proportion of effort, for all prior perspective and span conditions.

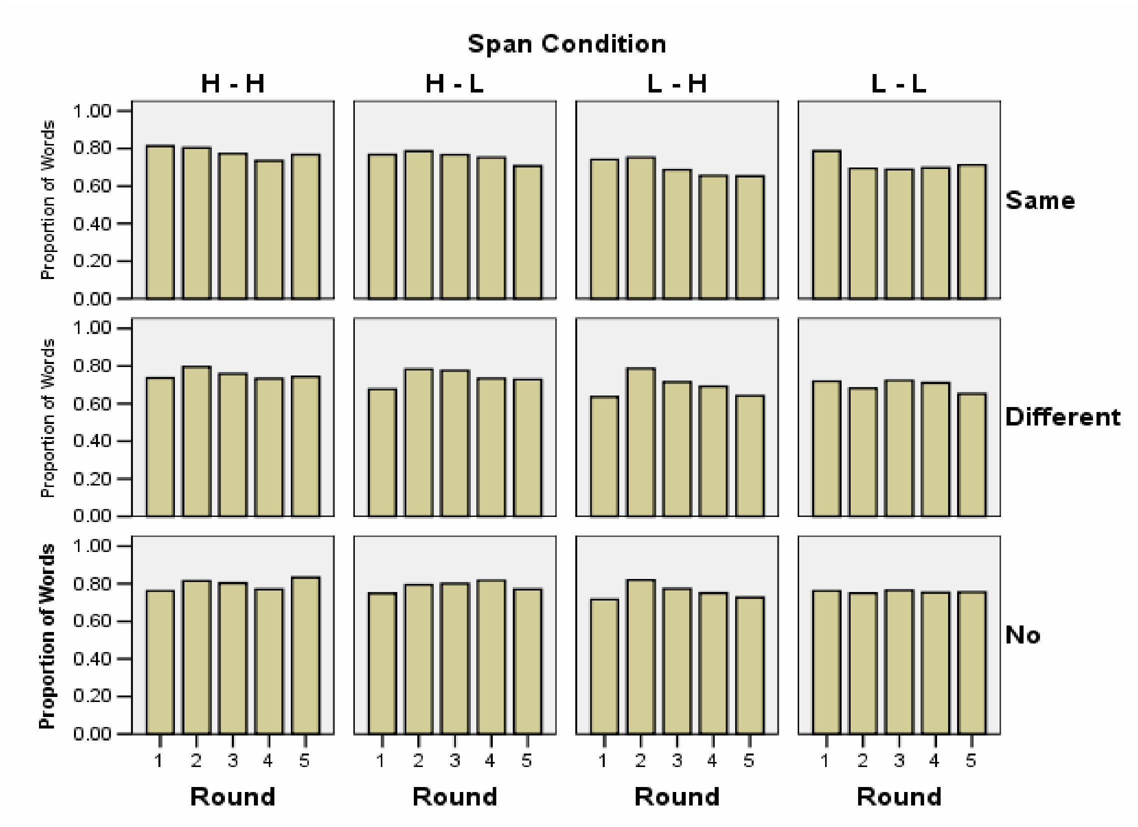


Figure 8a. Percentage of tangrams entrained: by round, for Different prior perspective condition, and for each span condition.

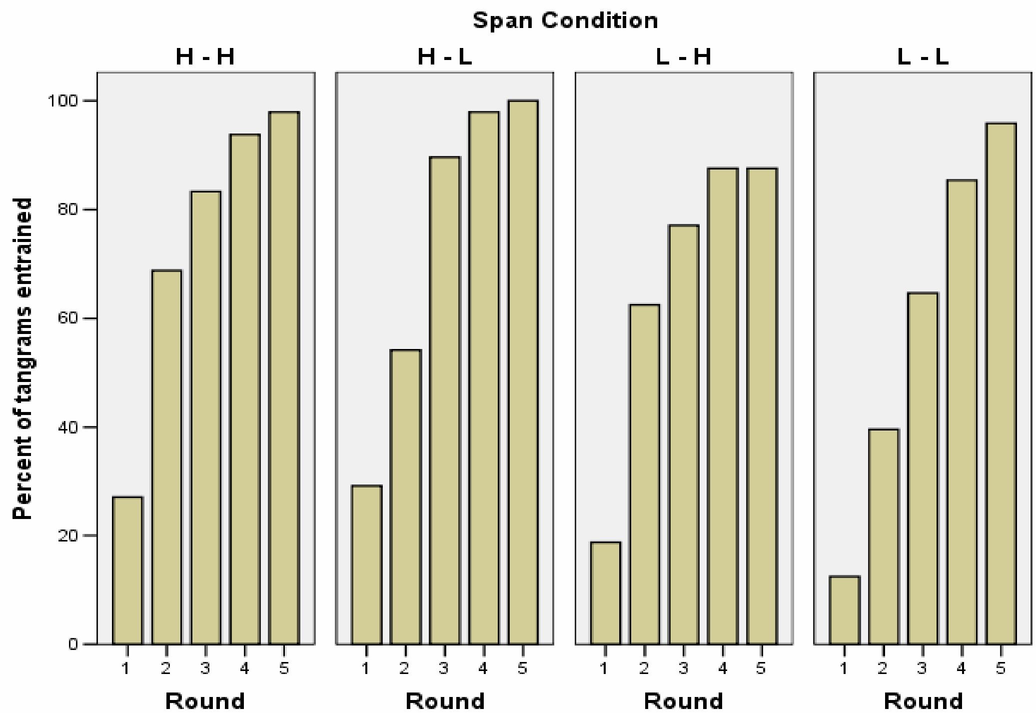


Figure 8b. Percentage of tangrams entrained by round: for Different prior perspective condition and for each span condition.

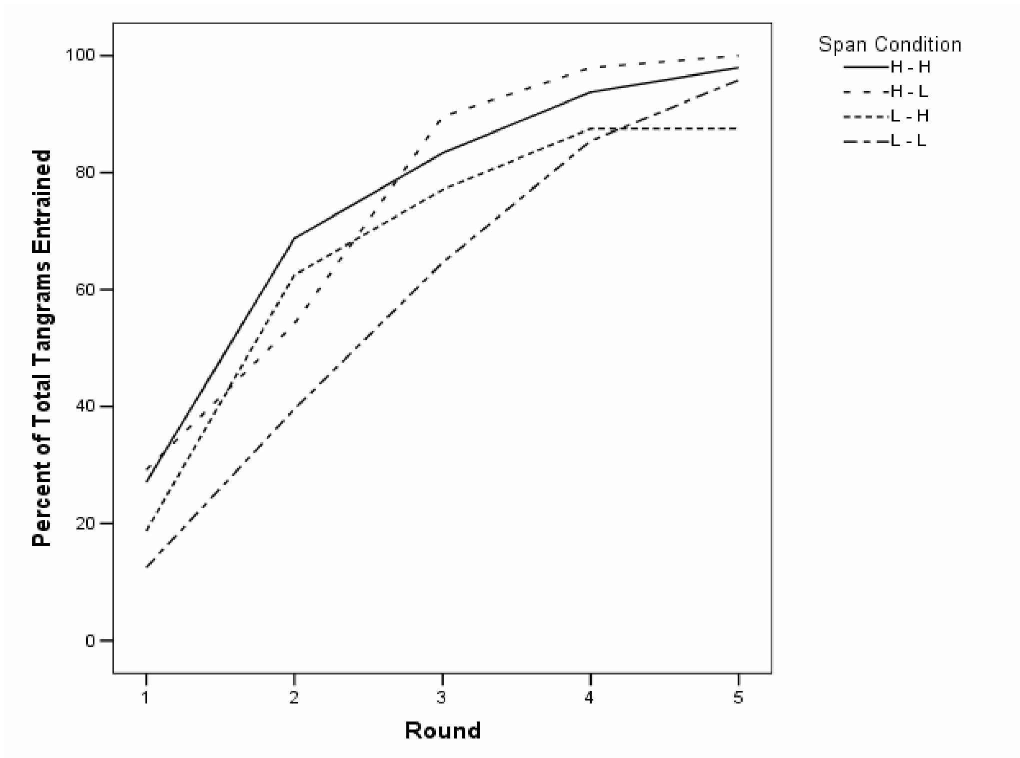


Figure 9. Percentage of tangrams where pairs ended up settling on the matcher's learned perspective: by round, for Different prior perspective condition and for each span condition.

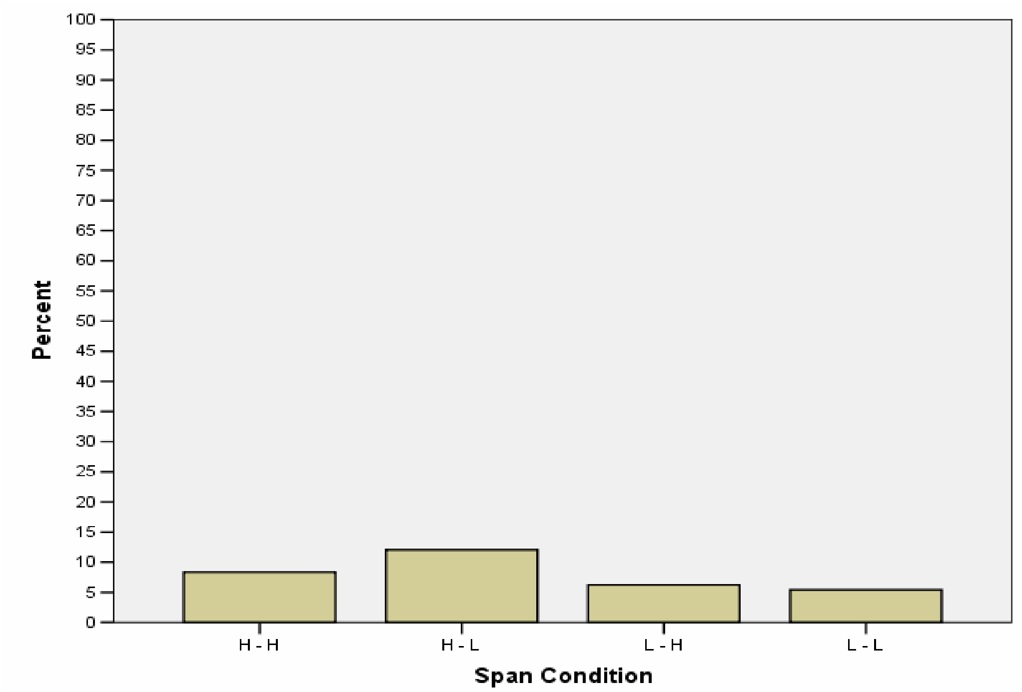


Figure 10. Percentage of tangrams where matchers offered counterproposals: by round, for the Different prior perspective condition and each span condition.

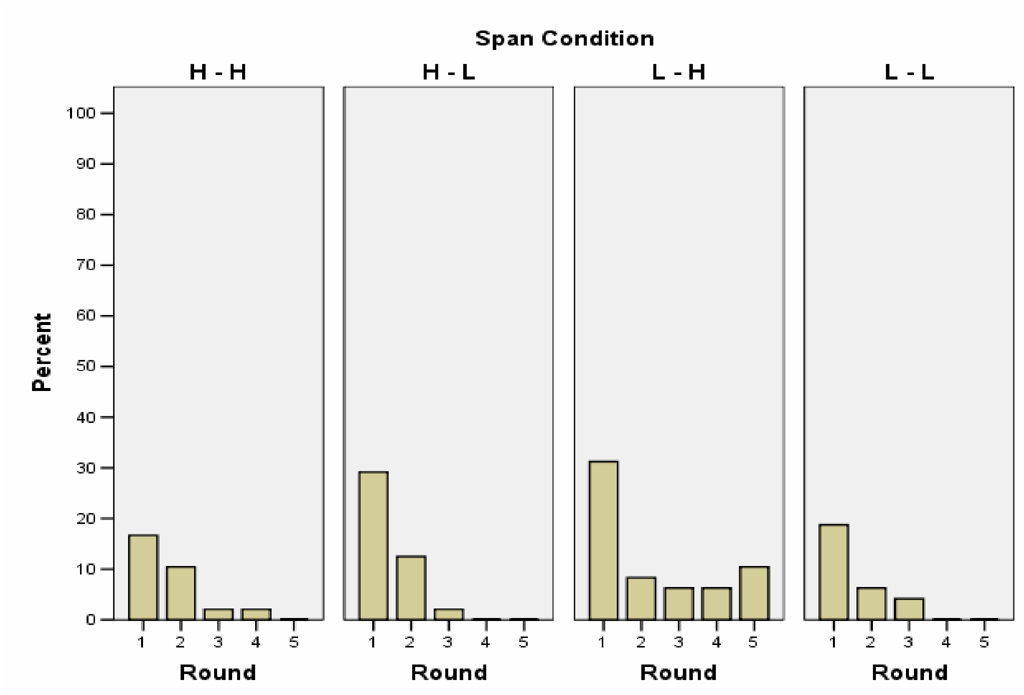


Figure 11. Proportion of matcher's counterproposals accepted by round, for each span condition.

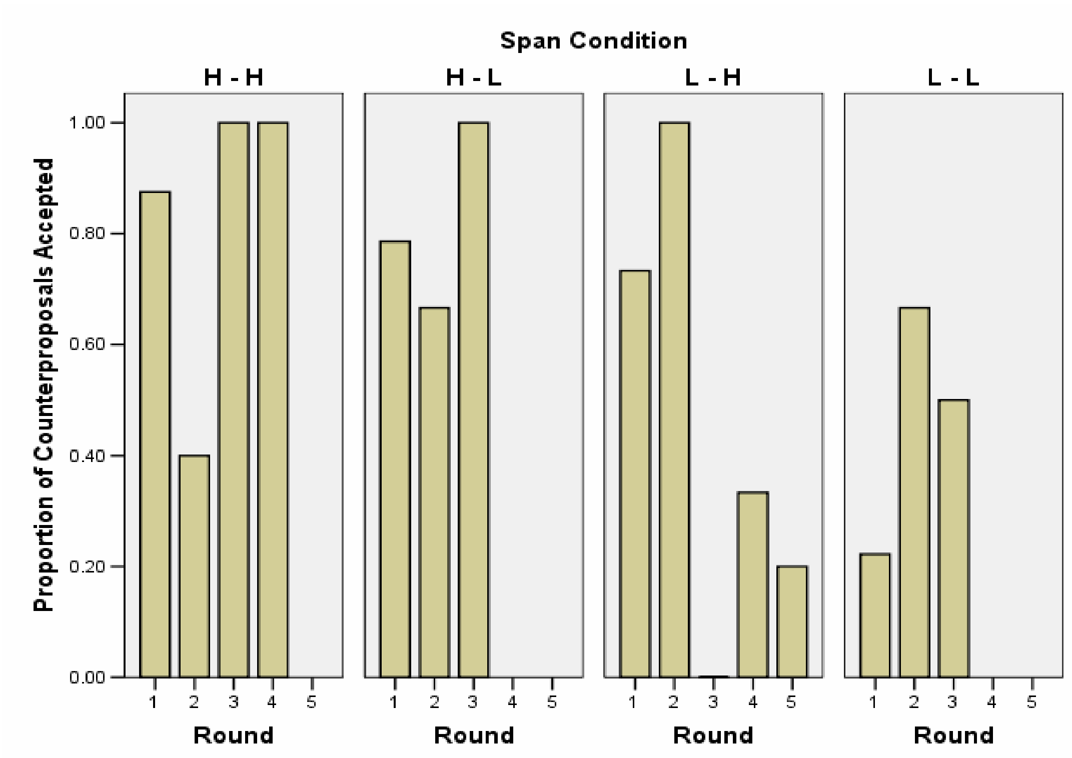


Figure 12. Percentage of tangrams where matchers requested clarification after the director's first turn.

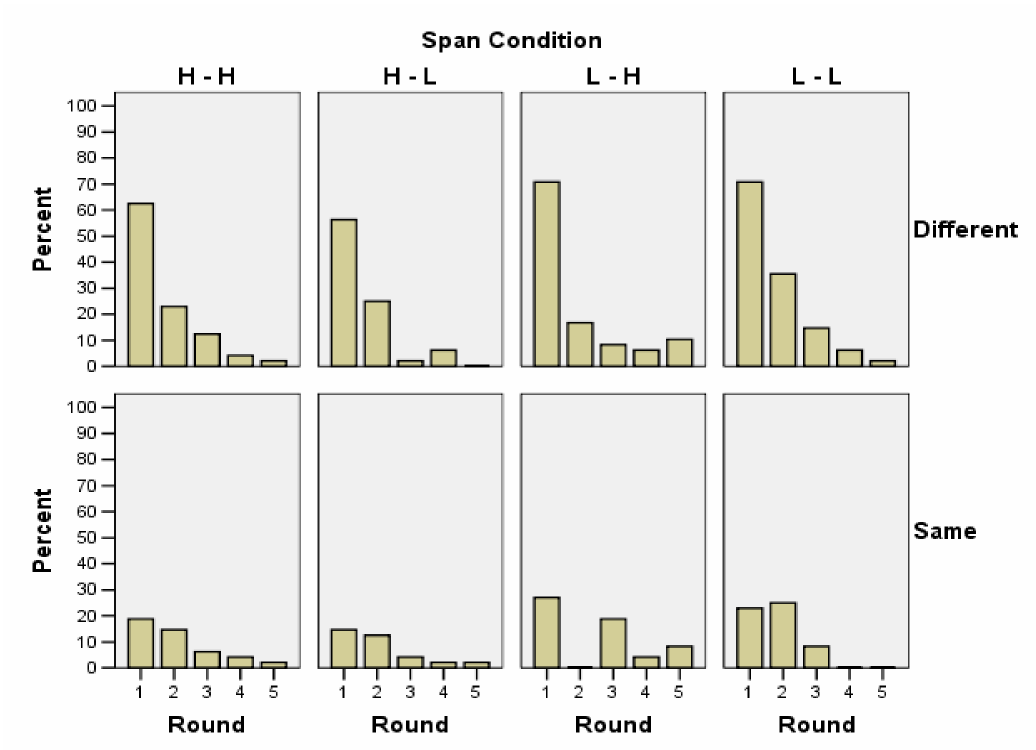


Figure 13. Percentage of the time that directors elaborated on their initial proposals: by round, for the Different and Same prior perspective conditions and each span condition.

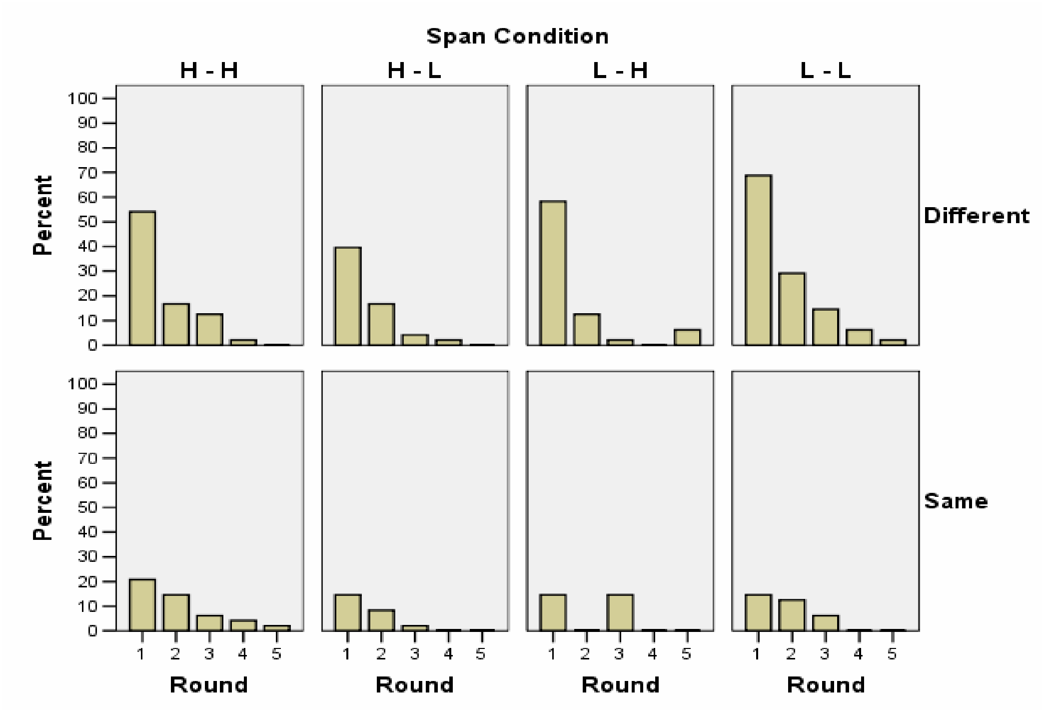


Figure 14. Percentage of tangrams where directors used a hybrid perspective on the 1st turn of a given round: for the Different prior perspective condition and each span condition.

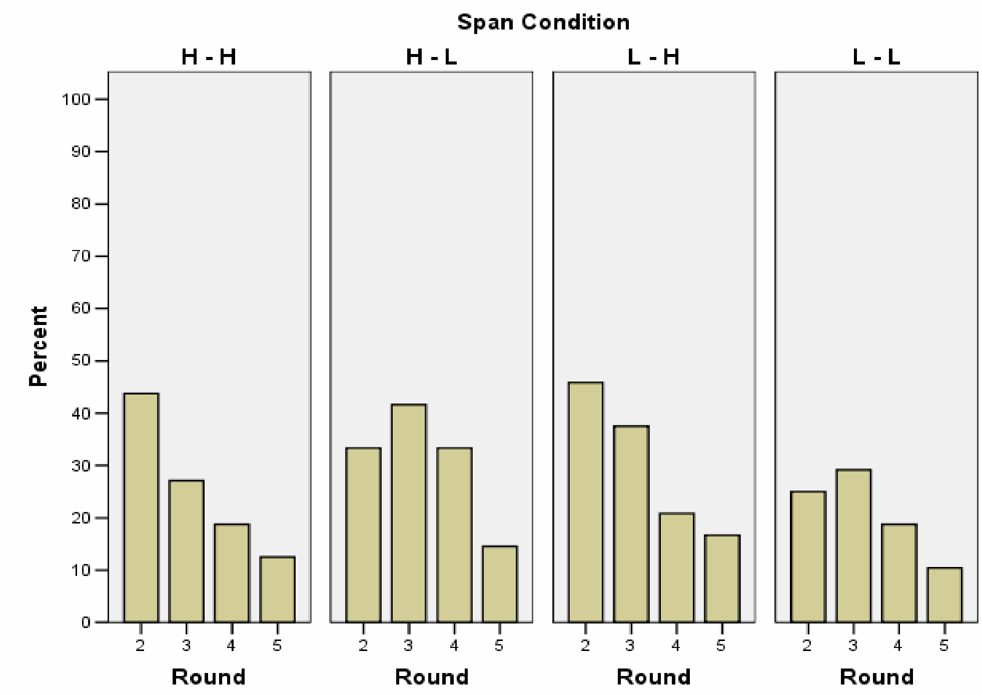


Figure 15a. Total mentions of the matcher’s learned perspective by round for all span conditions.

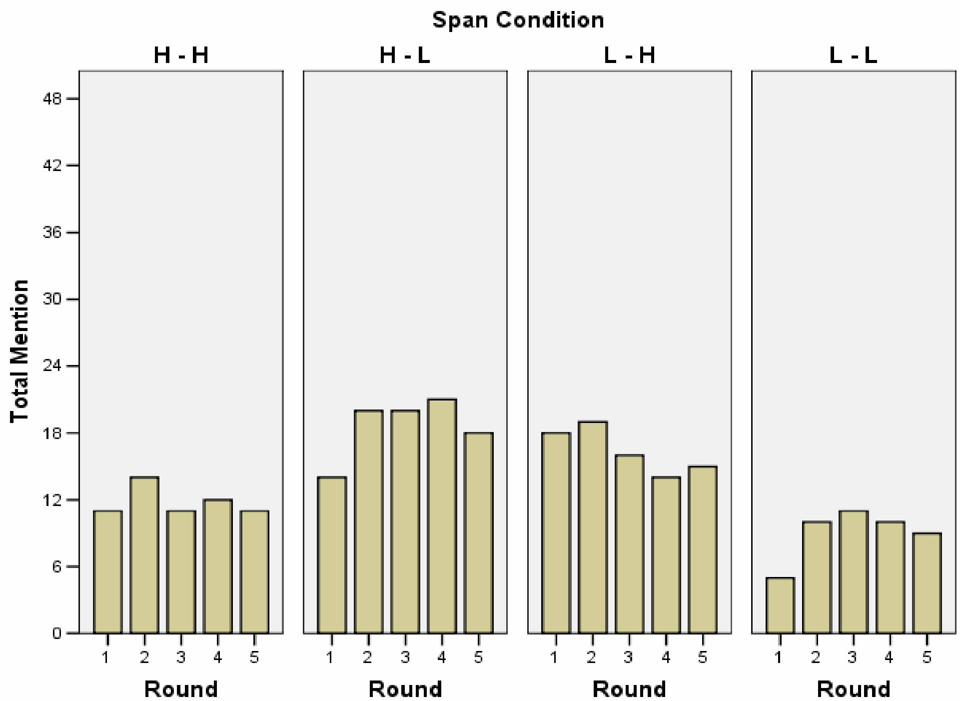


Figure 15b. When the matchers perspective was mentioned, who mentioned it:The number of times it was mentioned by the director or the matcher by round, for each span condition.

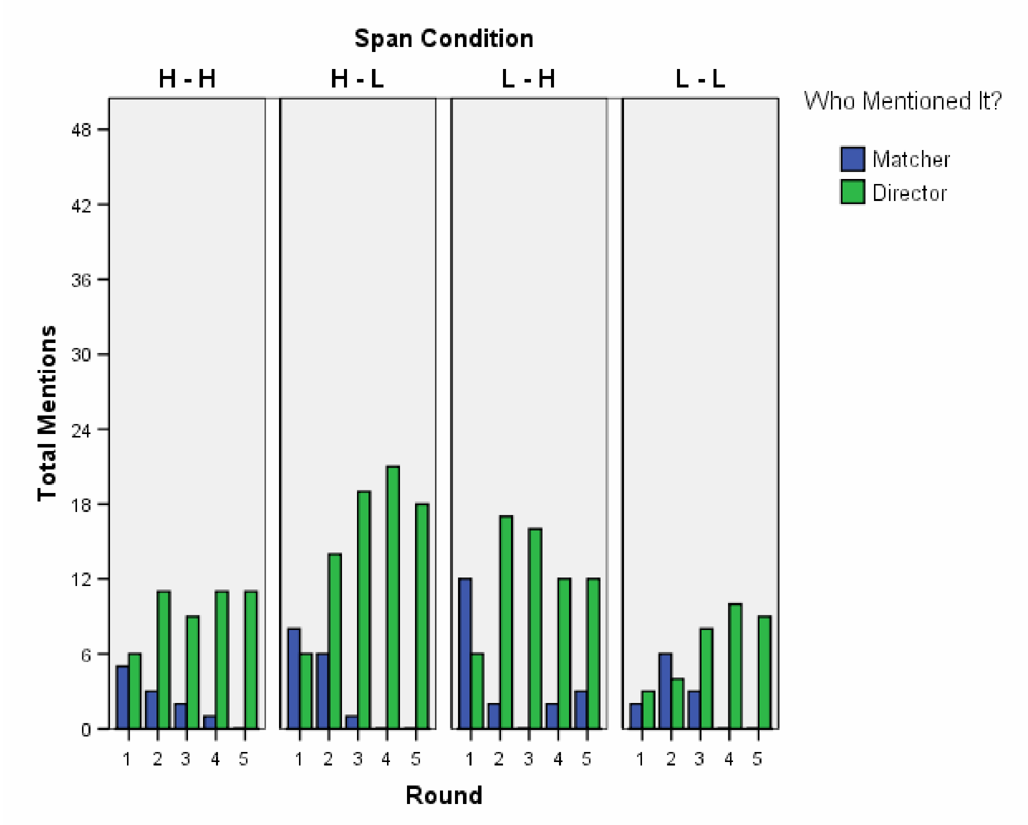


Table 1
Means of IRI scores for each span condition

	Span Condition							
	H_D-H_M		H_D-L_M		L_D-H_M		L_D-L_M	
IRI PT scale score	D	M	D	M	D	M	D	M
M	17.03	20.24	17.38	19.00	18.38	18.50	17.75	18.25
SD	5.00	3.36	3.81	2.40	3.28	3.39	4.63	3.13

Table 2
Balance of Gender for each pair of participants in Phase II by span condition

Gender Pair	Span Condition			
	H_D-H_M	H_D-L_M	L_D-H_M	L_D-L_M
Male director – Male matcher	2	1	1	2
Male director – Female matcher	3	3	1	2
Female director – Male matcher	1	2	1	2
Female matcher – Female matcher	2	2	5	2

Table 3
Mean word counts for each prior perspective and span condition

Prior Perspective Condition		Span Condition			
		H_D-H_M	H_D-L_M	L_D-H_M	L_D-L_M
Same	M	13.04	9.39	9.88	11.62
	SD	28.46	10.11	15.81	16.09
Different	M	21.13	14.52	17.3	22.23
	SD	49.77	21.65	32.45	37.19
No	M	23.79	22.73	22.25	38.28
	SD	39.56	28.9	36.19	75.56

Table 4

Percentage (in decimals) of tangrams entrained by round for all pairs in each span condition

Span Condition		Round 1	2	3	4	5
H_D-H_M	M	.60	.74	.88	.96	.97
	SD	.49	.44	.33	.20	.17
L_D-H_M	M	.38	.71	.79	.89	.90
	SD	.49	.46	.41	.32	.31
H_D-L_M	M	.46	.64	.89	.97	.98
	SD	.50	.48	.32	.17	.14
L_D-L_M	M	.31	.52	.74	.91	.96
	SD	.47	.50	.44	.29	.20

Appendix A. Entrainment Coding Categories.

1) Verbatim Perspective and Verbatim Phrase – The content words used in a round are identical (not necessarily in order) to those used in the immediately preceding round.

2) Shortening – The perspective is shortened in the following round; and the perspective in the following round must have appeared in the previous round *without* any new content words or phrases

3) Throwback – The perspective used in the round of interest is the same as, or contained within, one used in an earlier round, but not the immediately preceding one.

4) One or more content words different – The director either adds or alters one or more content words in a perspective between rounds

5) Some similarity – The perspective shares some similar content words with the previous round.

6) Entirely Different – The perspective used between rounds n and n - 1 are totally different, and n's perspective holds no similarity to anything previous

Appendix B. Perspective Content Coding Categories.

Single Perspectives

D	The director uses their learned perspective
M	Director uses the matcher's perspective
$\mathbf{N_D}$	Director uses a new perspective not previously learned
$\mathbf{N_M}$	Director uses a new perspective provided by the matcher not previously learned

Hybrids – (Combinations of one or more of the previous single perspective types into one phrase)

D +	The director's initial perspective is mentioned along with a some additional information
D M	A combination of both the director's and the matcher's learned perspectives
$\mathbf{D\ N_D}$	A combination of the director's learned perspective in addition to new perspective information provided by the director
$\mathbf{D\ N_M}$	A combination of the director's learned perspective in addition to new perspective information provided by the matcher
$\mathbf{D\ N_D N_M}$	A combination of the director's learned perspective in addition to new perspective information provided by both the director and matcher
$\mathbf{M\ N_D}$	A combination of the matcher's learned perspective in addition to new perspective information provided by the director
$\mathbf{M\ N_M}$	A combination of the matcher's learned perspective in addition to new perspective information also provided by the matcher
$\mathbf{M\ N_D\ N_M}$	A combination of the matcher's learned perspective in addition to new perspective information provided by both the director and matcher
$\mathbf{N_D\ N_M}$	A combination of new perspectives provided by both the director and matcher

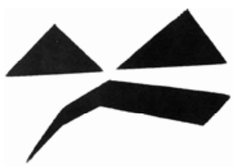
Appendix C. List of Tangram Stimuli with names.

Same Prior Perspective Condition

Swan

Movie Camera

Man raising hand



Jack o ’Lantern Face

Camel

Angel

Different Prior Perspective Condition

Dagger / Necktie

Cowboy Hat / Mountain Range

Barbell / Dog Bone

Arm making muscle/
Bird looking back

Arch / Legs

Viking Ship / Swimmer

No Prior Perspective Condition