THE UNIVERSITY OF CHICAGO

A THEORY OF DYNAMIC COORDINATION FOR CONVERSATIONAL INTERACTION

A DISSERTATION SUBMITTED TO THE FACULTY OF THE DIVISION OF THE SOCIAL SCIENCES IN CANDIDACY FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF PSYCHOLOGY

BY

DALE JEROME BARR, III

CHICAGO, ILLINOIS

DECEMBER 1999
TABLE OF CONTENTS

LIST OF FIGURES ............................................................................................................ v
LIST OF TABLES ............................................................................................................. vi
ACKNOWLEDGMENTS ................................................................................................ vii

CHAPTER ONE
COORDINATION AND CONVENTION IN DEFINITE REFERENCE
1.1 INTRODUCTION .................................................................................................. 1
1.2 DEFINITE REFERENCE ....................................................................................... 9
  1.2.2 Acts of Reference ........................................................................................ 12
  1.2.3 Definite Reference and Mutual Knowledge ................................................. 14
    1.2.3.1 Cognitive Implications of Mutual Knowledge ......................................... 18
1.3 CONVENTIONS AND REPEATED ACTS OF REFERENCE .......................... 20
1.4 THE THEORY OF DYNAMIC COORDINATION ............................................ 29
  1.4.1 Components of Dynamic Coordination ........................................................ 30
    1.4.1.1 Egocentric Anchoring and Adjustment ..................................................... 30
    1.4.1.2 Recycled Verbal Plans .............................................................................. 32
    1.4.1.3 Use of Prior Context ................................................................................. 33
    1.4.1.4 Evolution of Utterances ............................................................................ 36
    1.4.1.5 Multimodal Feedback Monitoring ............................................................ 38
  1.4.2 Implications and Predictions of the Theory .................................................. 39
    1.4.2.1 Lack of Partner-Specificity ....................................................................... 39
    1.4.2.2 The “Illusory Transparency of Convention” ............................................. 40

CHAPTER TWO
CONVENTION USE IN LANGUAGE PRODUCTION
2.1 INTRODUCTION ................................................................................................. 43
  2.1.1 The Use of Context in Message Generation................................................. 43
  2.1.2 Repeated Messages and the Evolution of Conventions ............................... 46
    2.1.2.1 Abbreviation ............................................................................................. 46
    2.1.2.2 Codification ............................................................................................... 49
  2.1.3 Dynamic Coordination in Convention Use ................................................. 52
    2.1.3.1 Abbreviation: Recycled Verbal Plans ....................................................... 52
    2.1.3.2 Codification: The Use of Prior Context .................................................... 53
2.2 METHODS ........................................................................................................... 54
  2.2.1 Task ............................................................................................................... 54
  2.2.2 Design ........................................................................................................... 56
  2.2.3 Procedure ...................................................................................................... 58
LIST OF TABLES

Table 2.1. Dependent Measures, Experiment One. ......................................................... 77
Table 3.1. Script for Descriptions of Nominal Filler Shapes......................................... 138
ACKNOWLEDGMENTS

The research reported in this dissertation was supported by Public Health Service grant R29 MH49685 to Boaz Keysar at the University of Chicago. Thanks to Kate Baldwin and Keith Murphy for helping to record the stimuli for Experiment Two. I would like to express my deepest gratitude to my advisor and committee chair, Boaz Keysar, for his mentorship and guidance over the course of my graduate training, and for providing an inspiring example as a scientist and as a human being. Thanks also to the members of my committee, David McNeill, Terry Regier, and Thomas Trabasso, for their valuable comments and encouragement. And finally, I thank my mother and father, Eileen and Dale Barr, whose love and support have made all of this possible.
CHAPTER ONE
COORDINATION AND CONVENTION IN DEFINITE REFERENCE

1.1 INTRODUCTION

In 1984, the semiotician Thomas Sebeok released a report that had been commissioned by the U. S. Office of Nuclear Waste Isolation. The Office of Nuclear Waste Isolation, charged with the disposal of radioactive materials, had come upon a problem that demanded the assistance of an expert in symbolic exchange. The problem was that radioactive waste has a life span that is longer than the recorded history of humankind. In the time that a ton of radioactive waste remains hazardous to human and other forms of terrestrial life, civilizations can rise and crumble, languages can be born and forgotten, libraries of scientific knowledge can turn to dust. How, these forethinking bureaucrats wondered, would it be possible to design a warning sign to inform future inhabitants of the locations of disposal sites? In other words, how would it be possible to generate a message whose meaning could be deciphered by cultures separated by millennia?

One solution that Sebeok came up with was to post the warning in all currently extant languages, in the hope that there was at least a single thread of cultural continuity such that the message could be deciphered. In this way, the commission could maximize the chances that future cultures could understand the message. Moreover,
assuming that these future individuals could decipher at least one of the many languages, this would give them the key to deciphering all of the other ancient and forgotten languages, such as English, in which the warning would be listed. It is amusing to imagine future government bureaucrats bringing the last living speaker of Guugu Yimithir to decipher this Rosetta stone, standing in the arid plains of Nevada, in front of an ancient toxic waste dump.

Apart from its science fiction potentials, this anecdote usefully illustrates several important facts about linguistic communication. Though the idea now seems a mere platitude, it has not always been recognized that the relationship between the form of a linguistic symbol and its meaning is fundamentally arbitrary. These arbitrary relationships are grounded in socially determined conventions. The word *moon* gains its meaning not through any inherent resemblance to a luminescent body in the night sky, but because there is a tacit agreement between speakers of English to associate this particular sequence of sounds with this concept. The key point is that conventions reflect social agreements rather than individual choices. If, on my own, I were to try to make the word *moon* refer to a side of beef, I would never be able to make this a reality, no matter how hard I tried.
Because of the arbitrary relationship between linguistic units and their meanings, language is only useful if interlocutors share knowledge of at least some of the governing conventions. Without this kind of intersubjectivity, symbolic\(^1\) exchange is impossible. Yet the problem is even harder, because over time, the conventional association between a word and its meaning can drift or radically change course. Etymologists track vagaries in the meanings of words over the course of generations. Conversation analysts can follow the subtle semantic drift of a word over the course of a single conversation.

Moreover, even when people agree on the meanings of words, they can still use them to mean different things. In different circumstances, the same words can be used to accomplish vastly diverse ends. A communicator can intend an utterance such as “you don’t know what real work is” to be construed in different ways. It could be construed as a literal statement of fact, an insult, or an ironic comment to a young doctoral candidate. The philosopher H. Paul Grice (1975) has shown that the divergence between the conventional, or literal meaning of a sentence and its meaning on a particular occasion of use, is itself governed by a set of conventions governing the flow of conversation. So not only are there conventions that determine the meaning of words, but conventions governing the use of conventions—second-order conventions that prescribe how sentences and words are to be used.

---

\(^{1}\) The word “symbolic” is used here in a Peircean sense; that is, a symbol is a kind of sign that is arbitrarily related to its referent. This is not to be confused with indices or icons, whose semiotic power has little to do with their conventionality. In cases where there can be little symbolic exchange, people can still use icons and indices to attempt to communicate, though it would be hard to successfully communicate a message about radiation sickness.
Language and language use are, therefore, shot through with conventions. As a consequence, language users face a vast amount of arbitrariness and uncertainty in language use, in every setting and at every time scale. Achieving intersubjectivity across distances is a problem that faces not only government bureaucrats looking for a way to send a message across millennia, but interlocutors in the temporal immediacy of face-to-face conversation. In consideration of the seeming intractable degree of arbitrariness in language, interlocutors seem to do extraordinarily well—our everyday experience tells us that in language use, misunderstanding is the exception, and understanding, the rule. If this is the case, then the average interlocutor must be equipped with some powerful techniques for perspective taking and ambiguity resolution. What are these techniques, and how are they used to achieve coordination? This is the subject of this thesis.

To state the question another way, what is the nature and locus of the “control mechanism” that governs how people coordinate meaning in language use? What makes it possible for people to effectively communicate ideas to one another when differences can be so vast? The elegance of coordination between conversational partners suggests the presence of centralized, intentional control mechanisms that guide the production and comprehension of individual utterances.

According to Herb Clark and colleagues, this control mechanism is common ground (the major works on this topic by Clark and colleagues are anthologized in Clark, 1992). Common ground is a special type of shared knowledge. A proposition \( p \) is assumed to be a part of the common ground between interlocutors A and B when they both know that \( p \), and know that they know that fact. Interlocutors can infer that a piece
of information is part of common ground based on a set of finite copresence heuristics. Coordination is maintained because communicators use these heuristics to adhere to common ground when they process language. When communicators design an utterance, they design it such that their addressee can understand it based on the meaning of the utterance along with their common ground (Clark & Marshall, 1981). Addressees, in turn, exploit this assumption to streamline how they process utterances (Clark & Carlson, 1981). When interlocutors strive to process information in a manner that is consistent with the knowledge that they share (and know that they share) with their partners, they maximize their chances for successful coordination at every turn.

Two questions arise as to the general assumptions of the theory of common ground. The first question is: As a theory of language processing, how realistic is the theory of common ground? The theory is more concerned with determining what kinds of psychological representations are necessary for interlocutors to successfully communicate than with how these representations might interact with the mechanisms governing the production and interpretation of individual utterances. It surveys conversation from a distant vantage point, far from the moment-by-moment psycholinguistic details of how actual utterances are produced and understood. How important are these details to the activity of coordinating meaning in language use?

As a cognitive activity, speaking and understanding are subject to constraints both on the kinds of representations that can be manipulated, and the speed at which these representations can be transformed. Comprehension and production must be accomplished in a finite amount of time, using a finite amount of resources. In the
terminology of artificial intelligence, it would seem that the pragmatic mechanisms that underlie coordination are more likely to consist of a set of “fast and dirty” heuristics than a set of exquisitely designed, specialized algorithms that guarantee communicative success, but are too complicated to be realized at the time-scale of conversation.

Previous work on on-line pragmatic processing, reviewed in detail in Chapters 2 and 3, suggest that the algorithms for computing common ground might be too complicated for interlocutors to effectively use at the time-scale of conversation (see Keysar, Barr, & Horton, 1998 and Keysar & Barr, in press, for reviews).

The second question regarding the theory of common ground is the degree to which it is even necessary for interlocutors to explicitly compute common ground to achieve coordination. The work of Clark and Marshall (1981) suggests that there would be no other alternative to this centralized, intentional control mechanism. Yet nature provides us with ample evidence of how other kinds of coordination can be achieved through the operation of very simple, distributed local mechanisms.

Consider the elegant behavior of many social aggregates of animals, from flocks of birds to schools of fish. When we observe a flock of birds burst into flight all at once or veer suddenly to avoid an obstacle, the flock seems to behave as a single unit. In cases such as these, we are tempted to assert the existence of some kind of centralized control mechanism that maintains the integrity of the flock at each decision point. How else could the birds stay together and avoid crashing into one another? For instance, we might assert that each individual bird is following a single “leader” bird, or is computing its average distance from all of the other birds in the flock. However, Reynolds (1987)
has shown that the behavior of the aggregate can emerge from local adjustments made by each bird. These adjustments are based on simple, plausible computations regarding each bird’s distance and speed relative to its closest neighbors—computations that are well within the grasp of an ordinary bird. When these small decisions are aggregated together in the entire flock, complex behavior emerges that is not directly evident from the joint operation of the local mechanisms.

What might this analogy say about theories of coordination in language use? One possible implication is that it may not be necessary for theories to build specialized coordination mechanisms into the heads of interlocutors. Perhaps the coordination that we observe in conversation emerges from the operation of simple cognitive principles, such as associative learning, concept formation, or how information is encoded and retrieved in memory. As Garrod and Anderson (1987) note, it would seem likely that language users possess “processes which take advantage of the inherently interactive and collaborative nature of dialogue to reduce the complexity of the inferences required for truly coordinated understanding in everyday conversation” (p. 185). Perhaps interlocutors do not explicitly compute common ground every time they formulate or interpret utterances—perhaps they compute it only when coordination fails. When interlocutors formulate or interpret utterances without considering common ground, there is some (currently unknown) likelihood that they will get it right anyway. Even if they do not, in conversation it is rare that they would only have one chance to do so. The inherently collaborative and multimodal nature of conversation affords the opportunity for interlocutors to continually monitor and incrementally adjust the level of coordination.
that they have achieved, like birds attempting to stay close to the neighboring birds in their flock.

This thesis constitutes an exploration of the possibility of locally based coordination in language use. I present a theory of linguistic coordination, called the “theory of dynamic coordination” to explain how interlocutors achieve coordination in conversational settings. Dynamic coordination theory postulates that interlocutors can achieve coordination through a set of very simple cognitive mechanisms that take advantage of the inherently interactive and multimodal settings of conversation. A principal contribution of the theory is that it clearly demonstrates that much of the coordination that occurs in conversations can be explained without recourse to intentional concepts, specialized operating principles like the principle of optimal design, or dedicated memory structures (i.e., Clark & Marshall’s “reference diaries”; Clark & Marshall, 1981).

To provide empirical support for the theory, I have conducted two experiments examining phenomena related to the emergence and use of conventions in repeated definite reference. The rest of this chapter provides the relevant background on theories of definite reference and convention use, and then outlines the theory of dynamic coordination. Two subsequent experiments provide empirical support. Chapter Two presents the first experiment, which examines convention use from the angle of utterance production. Chapter Three describes a second experiment, which investigates convention use from the opposite angle of utterance comprehension. Chapter Four summarizes the
results and discusses the larger implications of the theory and the experimental results for theories of language use.

1.2 DEFINITE REFERENCE

The phenomenon of definite reference provides an eminently fertile testing ground for theories of language use, because the principles governing successful acts of reference have been clearly articulated in previous work by philosophers, linguists, and psychologists. In modern research on language use, the theory of mutual knowledge (or common ground) was developed specifically to account for how communicators perform successful acts of reference.

Communicators are said to “refer” when they use some description to “pick out” a specific entity or set of entities as the intersubjective focus of attention. This entity or set of entities is known as the referent. Communicators usually refer to some referent when they wish to predicate some property of it. Thus, when your friend Sheila tells you, “the hitchhiker I picked up yesterday was an Argentinean writer”, the phrase “the hitchhiker I picked up yesterday” establishes a referent for the predication “was an Argentinean writer”. This kind of phrase is known as a definite description. As in Sheila’s description, most definite descriptions involve the definite article “the” and some phrase that describes the referent.² Once you locate the referent of Sheila’s description,

² Because I favor a functional rather than formal definition of definite description, certain phrases are still definite descriptions even when communicators elide the definite article “the” (as many participants in Experiment 1 did).
you could then update your knowledge about the referent by adding the feature “was an Argentinean writer”.

When your friend uses the description “the hitchhiker I picked up yesterday”, the use of the phrase carries several presuppositions. How these presuppositions are defined varies from scholar to scholar (see, for instance, Russell, 1920; Searle, 1969; Hawkins, 1968; Clark & Marshall, 1981). I have attempted to distill this work into a set of three: the presupposition of existence, uniqueness, and identifiability.

First, when Sheila says “the hitchhiker I picked up yesterday”, she presupposes that the entity which fits the description “the hitchhiker I picked up yesterday” actually exists (Russell, 1920). This can be referred to as the presupposition of existence. This presupposition is somewhat trivial and tautological, in the sense that anything that can be referred to can be said to exist in some sense, even when its existence is fictional; as is the case for unicorns, golden mountains, and round squares, which exist at least as ideas (Russell, 1920; Searle, 1969). It would be anomalous for Sheila to use the phrase “the hitchhiker I picked up yesterday” if, in fact, there were no real or imaginary referent that fit this description. If we somehow knew that this presupposition had not been met, we would probably think that Sheila was lying or was suffering from one of her famous lapses of memory.

Furthermore, Sheila presupposes that there exists one and only one referent that satisfies the description “the hitchhiker I picked up yesterday”. This can be termed the presupposition of uniqueness. If Sheila picked up two hitchhikers yesterday, then her use of the phrase “the hitchhiker I picked up yesterday” would be inadequate, because it
would fail to uniquely pick out a referent. In this case, we would probably ask Sheila, “which one?” In Searle’s terms, the act of reference would not be “fully consummated”—a unique referent would not be intersubjectively established—until this presupposition was satisfied.3

Finally, Sheila’s description assumes you already know that she picked up a hitchhiker, or that you can readily infer that she did. In other words, she presupposes that you will be able to identify the referent from within a shared set of objects (Hawkins, 1968; Chafe, 1976) or, at least, infer its unique existence on some basis (Clark & Marshall, 1981). This is the presupposition of identifiability. This last presupposition differentiates the use of definite description from another class of description with a referring function, known as indefinite description.

Unlike definite descriptions, communicators use indefinite descriptions when the referent of the description would not meet the presupposition of identifiability. Indefinite descriptions are typically used with the indefinite article, “a”, as when Sheila tells you “I picked up a hitchhiker again yesterday”. The phrase “a hitchhiker” tells you something about the kind of individual that she picked up, but it does not presuppose that you, the hearer, will be able to identify a particular individual as a referent (outside of the trivial fact that it establishes an single entity “hitchhiker” within the reference frame of Sheila’s storytelling; see Strawson, 1959).

3 The presupposition of uniqueness only holds for descriptions that are not “attributive”.
There is a good reason why it is preferable to define definite descriptions using functional rather than formal criteria. This is because descriptions such as “the hitchhiker” have not only referential, but also attributive uses (Donnellan, 1966). Imagine that Sheila has picked up another hitchhiker, and you visit her as she recovers from amnesia in a local hospital. You tell her that “the hitchhiker you picked up yesterday, whoever it was, robbed you clean”. In this case, the individual that you have in mind is anyone who fits the description “the hitchhiker you picked up yesterday”. Because of this, it is more like an indefinite description than a definite description. Accordingly, the studies in this thesis do not concern attributive descriptions.

1.2.2 Acts of Reference

Referents can be described in various ways. For instance, a rose can be described as “the thing in the vase”, “the flower”, “the red flower”, or “the rose”, even though by any description, it still smells as sweet. However, in a given instance of reference, not just any description will satisfy the presuppositions of uniqueness and identifiability. Many possible descriptions would therefore fail to consummate reference. How, then, do communicators decide how to describe a referent?

When formulating a definite description, a communicator performs a kind of speech act (Austin, 1962; Searle, 1969). As in other speech acts, communicators strive to make sure that the relevant “preparatory conditions” have been met—in the case of definite reference, the presuppositions of existence, uniqueness, and identifiability. When an act of reference meets these standards, it can be called felicitous.
Since Grice (1975), the selection of information to be included in a definite description has been assumed to be governed by the cooperative principle. The cooperative principle requires communicators to “make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged” (p. 45). More specifically, communicators who are cooperative should observe the Maxim of Quantity, which states that they should provide sufficient, but not excessive, information for the hearer to be able to identify the referent uniquely—i.e., in order to make sure that the presuppositions of identifiability and uniqueness are fulfilled. Communicators, like the fabled Goldilocks, must seek out descriptions which are neither too hot (overinformative), nor too cold (underinformative), but just right.

Communicators should take care not to violate the Maxim of Quantity when they formulate descriptions. Otherwise, their referential acts can “misfire”, often with unintended conversational side effects. If a communicator insufficiently describes a referent, then the addressee will fail to find a unique referent that fits the description. For instance, the use of the description “the boy with the nose” to pick out a referent would probably always be infelicitous, because it fails to satisfy the presupposition of uniqueness—it cannot pick out a single boy from among many. An addressee who hears this description will assume that the communicator is being cooperative and that they have therefore provided enough information in the description for them to uniquely identify the referent. Rather than abandon this assumption, when they hear an underinformative description such as “the boy with the nose”, they will likely infer that
the boy’s nose must have some abnormal property (e.g., it is unusually large or made of porcelain) that would enable unique identification of the particular boy. In other words, to preserve the assumption that the communicator is being cooperative, they derive an “implicature” that was not directly communicated by the utterance. The addressee may go even further and infer that the communicator is withholding information about the abnormal characteristics of the boy’s nose in order to be polite.

For these reasons, if communicators want to make sure that the conversational implicatures (Grice, 1975) and perlocutionary effects (Austin, 1962) of their utterances are predictable, then they should be sure to make sure their descriptions are “just right”. However, communicators are much less fortunate than Goldilocks, who had only three porridges to choose from—communicators must choose from among an infinite set of porridges. How do communicators make sure the ones they choose are “just right” for their interlocutor?

### 1.2.3 Definite Reference and Mutual Knowledge

In an influential paper, Clark and Marshall (1981) addressed this problem by introducing another presupposition—mutual knowledge—that is assumed to govern felicitous acts of definite reference. Since this important paper, the theory of mutual knowledge or common ground has been advanced as a general theory of language use (Clark, 1992). Their basic insight was that referential communication is a kind of coordination problem (Schelling, 1960).
Coordination problems have a special kind of mutual dependency structure that differentiates them from other kinds of problems. In a coordination problem, the outcome of an alternative depends upon the choices that other people make. For instance, imagine that you have become separated from your friend at the airport. You would both be thrust into the coordination problem of finding one another. Your decision of whether to walk around or stand and wait would depend upon what you expect your friend to do. The optimal solution is usually for one person to sit and wait while the other person goes wandering. Yet who is to sit, and who is to search? To decide on a plan of action, you would have to consider what you expect your friend to do, and what your friend would expect you to do, and what you expect your friend to expect you to expect your friend to do, etc. This kind of reasoning can iterate endlessly, and all hope would be lost if both you and your friend decided to sit and try to reason the problem out before making a decision.4

In what sense is referential communication a kind of coordination problem? Clark and Marshall argue that whether a referential act is felicitous or not depends upon what the communicator and interlocutor mutually know. They start by showing that acts of reference based on shared knowledge alone are not guaranteed to be successful. When Ann asks Bob, “Have you seen the movie playing at the Roxy tonight?”, how would Ann be sure that Bob would be able to identify the referent of “the movie playing at the Roxy tonight”? If they both know that the movie is A Day at the Races, but Ann

---

4 For this reason, philosophers and scholars of language use should stay clear of airports.
does not know that Bob knows this fact, then Ann might perform an infelicitous act of reference, because her utterance would not be guaranteed to fulfill the presupposition of identifiability.

Therefore, not only must they both know which movie is at the Roxy, but Ann must know that Bob knows this fact. But things get even more complicated very quickly. For the sake of simplicity, let us follow Clark and Marshall and assign a variable $p$ to the proposition that *A Day at the Races* is playing at the Roxy. Clark and Marshall go on to show that it is insufficient to add the condition that Ann knows that Bob knows that $p$, because even when Ann knows that this additional condition has been satisfied, Ann could not be sure that Bob knows that Ann knows this fact. Perhaps Bob noted in the morning paper that the movie had been changed at the last minute from *Monkey Business* to *A Day at the Races*, but he can’t be sure that Ann knows that he knows that fact. In this case, he might think that Ann would be referring to *Monkey Business*. So for Ann’s act of reference to be felicitous, an extra condition—that she knows that Bob knows that she knows that Bob knows that $p$—would need to be added.

This kind of recursion can be carried out indefinitely, and in each case, it is possible to think up clever scenarios whereby Ann’s description would fail to satisfy the presupposition of identifiability and therefore would be infelicitous. For Ann’s speech act to be felicitous, it would be necessary for Ann and Bob to *mutually know* that $p$—they would have to know that $p$ and know that they know that fact.

This exercise is the scenic route that leads into the philosophical quagmire known as the *mutual knowledge paradox*. For Ann to perform a felicitous act of
reference, she would have to be sure that she and Bob mutually know that $p$. But for Ann to know this, she would have to verify an infinite number of conditions before performing her act. Yet referential acts are performed in a finite amount of time. How can this paradox be resolved?

There are two roads leading out of the quagmire. The first road asserts that mutual knowledge is a psychological impossibility, and because of this, it cannot form part of a theory of language use (Sperber and Wilson, 1982, 1986/1995). This implies that communicators do not fully check all of the conditions that determine whether their referential acts are felicitous or not, but simply take a chance that their acts might fail. The other road out attempts to preserve the assumption that communicators strive to be felicitous by assuming that they can infer that the infinity of conditions have been satisfied using finite means. In other words, they do not have to laboriously check each condition, but on the basis of certain kinds of evidence, they can infer all at once that these conditions have been satisfied. This is the road that Clark and Marshall chose. They argued that addressees have at their disposal a set of three finite heuristics for inferring mutual knowledge, which they called copresence heuristics: physical copresence, linguistic copresence, and community membership.

Communicators can use the knowledge that something is copresent via one of these heuristics to infer that it is also mutually known. The heuristics differ in the sources of evidence that are used to infer mutual knowledge. The physical copresence heuristic states that anything that is perceptually available to both Ann and Bob can be inferred as mutually known. Imagine that Ann and Bob stand in front of a large poster
that says “Now Playing at the Roxy. The Marx Brothers in: A Day at the Races”. Because this poster is physically copresent with Bob, Ann can infer that Bob would be able to uniquely identify the referent of the phrase “the movie at the Roxy”.

Linguistic copresence works in a similar manner. When a piece of information has been established as part of the discourse record between Ann and Bob, then they can assume mutual knowledge of it. If in an earlier conversation, Carla informed Ann that p in the presence of Bob, Ann and Bob could infer mutual knowledge of p. Finally, if they know that they are both members of a community that would have the information—e.g., movie fanatics who religiously read the local listings—they could infer mutual knowledge on the basis of community membership.

1.2.3.1 Cognitive Implications of Mutual Knowledge

If, as Clark and Marshall claim, mutual knowledge is a routine part of language use, then this fact should have implications for how information is stored in memory and accessed during language use.

Organization of Memory. According to Clark and Marshall, people maintain specialized memory structures known as “reference diaries” to keep track of their common ground with others (Clark & Marshall, 1981). Not only do people remember facts, but they also remember with whom these facts are mutually known. For example, when Ann and Bob wave at each other as they leave the cinema after viewing the film Annie Hall, not only will Ann encode the information that she and Bob both saw the movie, and that they greeted one another there, but that they mutually know these two
facts. Two days later, when Bob calls Ann and asks, “What did you think of the movie?”, Ann should consult her reference diary for Bob to find the appropriate referent. On this basis, she would understand that Bob is referring to *Annie Hall* and not *Blow Up*, which she had seen the night before with Carla.

**Optimal Design.** The claim that people manage information in the form of reference diaries makes specific empirical predictions about memory encoding and retrieval during language use. Reference diaries provide a knowledge base from which people formulate and interpret messages with respect to their specific interlocutors. When communicators design an utterance for a specific addressee, they should use only the mutually known information included in the reference diary for that person. This assumption is embodied in the *principle of optimal design*, which states that a communicator “designs his utterance in such a way that he has good reason to believe that the addressee can readily and uniquely compute what he meant on the basis of the utterance along with the rest of their common ground.” (Clark, Schreuder, & Buttrick, 1983, p. 246). Optimal design is also assumed to play a role in message comprehension, because addressees expect that communicators obey this principle. The assumption is that addressees would understand a message optimally if they exclusively access the common ground information stored in their reference diary for the communicator: “...the
comprehension process must keep track of common ground, and its performance will be optimal if it limits its access to that common ground.” (Clark & Carlson, 1981, p. 328).

1.3 CONVENTIONS AND REPEATED ACTS OF REFERENCE

When communicators and addressees establish reference over repeated turns, they establish conventions for how referents are to be described. In this section, I begin with a brief definition of convention as formulated by the philosopher Lewis (1968). I outline the implications of the theory of common ground for how conventions are used in language production and comprehension, saving specific details about language processing for Chapters 2 and 3. Next, I propose a new theory—the theory of dynamic coordination—that explains convention use through the application of simple cognitive mechanisms, but does not invoke mutual knowledge. I discuss this theory in detail at the end of the chapter.

In Lewis’s analysis (1968), conventions constitute essentially arbitrary solutions to recurring coordination problems. Consider the convention governing the flow of pedestrian traffic in American cities and towns. Modern pedestrians face the problem of diminishing sidewalk space, as civic planners perpetuate the hegemony of the automobile by designing larger roads with smaller sidewalks. Because space is at a premium, pedestrians face a recurring problem of avoiding head-on collisions. To deal with this problem, pedestrians in the United States have established the convention that

---

5 There is an important ambiguity in the way that Clark & Carlson express their prediction that the comprehension process “limits its access to that common ground”. From the context of the citation it is clear that they intend this to mean that the comprehension process considers common ground exclusively, and not in the sense that its access to common ground information should be limited.
oncoming pedestrians should pass on the right. That this convention is arbitrary is made clear by the fact that in England, the convention is to pass on the left (or so I am told).

As with other conventions, the evolution of this solution would have followed a “salience-precedence-convention” path outlined by Lewis (1968). To elucidate the details of Lewis’s theory, here is a completely apocryphal “folk history” of how the pedestrian traffic convention may have arisen.

Historically, the traffic patterns of pedestrians were not confined to the small strips of concrete that form the modern sidewalk. From colonial days to the advent of the automobile, pedestrians could wander to their destinations through the streets of the city using the full breadth of the road, easily scuttling away from the occasional slow-moving horse and buggy. Pictures from olden days confirm the impression of a disorder of movement. In that age, ample space for walking ensured that head-on collisions between pedestrians were extremely rare, and because it was not a recurring problem, no convention was needed. Once automobiles arrived on the scene, however, the booming auto industry and safety concerns demanded that pedestrian traffic be confined to smaller and smaller areas, bringing the now ubiquitous city sidewalk into being.

This shrinking of walking space had the result that pedestrians faced a recurring problem of head-on collisions. This problem was effectively solved by establishing the convention to pass on the right. As with most other conventions, this was probably not decreed from the bell tower, but emerged from the grassroots. Let us imagine the momentous occasion when the first head-on collision was avoided by passing on the right. When these two pioneers approached one another, for whatever reason, it
was most salient for them to pass on the right, perhaps because it required a smaller adjustment in their trajectories than would passing on the left. This event established a precedent that became more and more convention-like as these two individuals passed each other each morning. They propagated this precedent to other people, who in turn, followed the same precedent to avoid subsequent encounters with other citizens, and so on. Over time, the precedent spread to the entire population, eventually becoming a mutually known convention, and the young nation rejoiced as many skinned knees and bruised elbows were avoided.

What differentiates a convention from a precedent? According to Lewis (1968), the hallmark of a convention is that it is chosen as a course of action even when other alternatives are more salient. To illustrate, let us bring our folk history up to the present day. Imagine that you are walking southbound across the Michigan Avenue Bridge in Chicago, and find yourself on a collision course with another pedestrian walking northbound. You approach the other pedestrian at such an angle that you would bump right shoulders as you passed on the left. Based on your current trajectory, the easiest and most salient alternative would be to simply adjust slightly to your left when passing, thereby continuing on your course while leaving the pedestrian unharmed. A less salient possibility would be to make a larger shift to the right, which would bring you in line with the pedestrian traffic code. However, this alternative is less desirable for you, because it would require you to shift the Crate & Barrel shopping bag you carry in your left hand to your right. If you did not shift the bag, which is over-laden with gourmet cooking items, it would be in danger of slamming your potentially litigious and medically
uninsured fellow pedestrian in the kneecap. Though it would be simpler, and more
salient, to pass on your left, you can’t be sure that the other pedestrian, who is currently
enjoying the view of the Tribune Tower, would not try to follow the convention and
suddenly plan to a shift to the right once she became cognizant of the situation. If this
happened, your hurried pace would send you into a full-frontal, Monday-night-football
style collision. Your best guess is that she would follow the convention to pass on the
right, and therefore you effortfully hoist your bag across your body from left to right and
cross to the other side of the sidewalk. You decide that this would be the best way to
avoid ending up in a tumble of broken banana hangers and scattered grounds of
Colombian Supremo, and would at least spare you the embarrassment of performing an
intimate collision-avoidance dance with a total stranger while you negotiated a clear path.
Though you choose an alternative in line with the convention, it is neither the least
effortful nor the most salient, but it seems the most likely to succeed.

In language, as in pedestrian traffic, conventions do not always minimize
effort, nor are they necessarily the most salient alternatives. The English language has
over two hundred verbs that form the past tense irregularly, making it effortful for a non-
Native speaker to learn and use correctly. It might be easiest and most salient for a
person learning English to just add –ed to everything—after all, people would probably
still understand the meaning. However, English speakers, perhaps still harboring a vague
feeling of resentment from their own struggles as a child to avoid overregularizing the
past tense, will simply not tolerate people walking around saying breaked and dranked.
Alas, one has no choice but to acquiesce to the reign of convention.
The existence of precedents and conventions can drastically simplify the problem of formulating definite descriptions. Though communicators can describe a referent in an infinite number of ways, the existence of a precedent can collapse that infinite space down to a single point. Let us return to the case of Ann and Bob. Suppose that Ann finds out that Bob has seen Woody Allen’s film *Sleeper*, now playing at the Roxy. She wants to let Bob know that she especially liked a particular scene, so she must come up with a definite description to refer to it. She asks Bob, “Do you remember the really funny scene where Woody Allen peels an enormous banana and then slips on the peel *a la* Harpo Marx? That was my favorite scene”. Bob, a pretentious music snob who collects movie soundtracks, typically does not find such paltry comic devices very memorable. He remembers the banana peel scene only because it was accompanied by a soundtrack of Dixieland jazz performed by the Preservation Hall Jazz Band, notable because it featured Woody himself on clarinet. Because the soundtrack was more salient to him than the slapstick humor, he would have chosen a different definite description for the scene, like “the scene where Woody plays clarinet on the soundtrack”. Even though this description would be more salient for him, when he refers to the scene in the future, he might acquiesce to Ann’s precedent and call it “the banana scene”, thereby transforming her description into a full-fledged convention.

In the theory of mutual knowledge, it is the belief about what others know that guides the production and interpretation of definite descriptions such as Ann’s initial description of the scene. Following Lewis (1968), Clark and Carlson (1981) argue that mutual knowledge is essential for the use of conventions. They point out that “the source
of common ground for conventions... is community membership.” (p. 73). When Ann attempts to describe the scene to Bob, she should select a description that would be salient to both of them with respect to their mutual knowledge. She can infer certain things as mutually known with Bob by virtue of their joint membership in certain communities. Her description is initially rather lengthy because she cannot assume that a description such as “the funny banana peel scene” would be readily understood by Bob, especially if she takes into account that Bob is not a member of the community of people who appreciate the sophistication of certain kinds of meta-humor, such as Woody Allen’s self-conscious recycling of an old comic cliché.

As Ann continues to talk about the scene with Bob, each time she refers to it, her description gets shorter and more conventionalized, to the point that she describes it simply as “the banana scene”. But Bob has ceased to be a receptive listener, so she decides to share her appreciation of the scene with her other friends, starting with Carla and proceeding through her address book to Z. This morning, Ann found out from Carla’s boyfriend David that Carla and David had gone to see *Sleeper* at the Roxy last night. She decides to call up Carla with the pretense of organizing a fondue party for the weekend. It is not yet mutually known between Ann and Carla that Carla has seen the movie. But Ann’s need to share her obsession for this scene is deep and immediate. The question is, how will Ann describe the scene to Carla? One possibility is that she will produce an elaborate description, perhaps as long as her initial description of the scene when she described it to Bob. Mutual knowledge predicts that differences in the overlap of communities that Carla and Ann belonged to, relative to Ann’s overlap with Bob,
would lead Ann to formulate a description which would differ in important ways from her earlier descriptions to Bob. On the other hand, the convention that she established with Bob is very salient to her, because she was just using it in her conversation with Bob. However, if she were to use this convention with Carla, she would be egocentric and uncooperative.

What does Ann do? Let us assume that Carla picks up the phone, and the first thing that Ann says to her is, “So... how did you like the banana scene?” Note how natural this description seems to be, even though it outrightly appears to violate the standards set by mutual knowledge and cooperativeness. Some theorists’ intuitions tell us—as mine do—that descriptions such as these are normal, even routine, in everyday conversation, even though they might violate mutual knowledge. Keysar (1994) would classify this kind of description as demonstrating an “illusory transparency of intention”—that is, as showing that Ann’s egocentric belief that the intended referent of her phrase would be immediately clear to Carla. Perhaps Carla will think ‘The Banana Scene’ is the name of a new progressive rock group.

However, theorists may have other intuitions about Ann’s description in this case. Ann should only describe the scene as “the banana scene” if she has good reason to believe that Carla would be able to immediately identify the referent based on their mutual knowledge (Clark & Marshall, 1981). In other words, she should follow the principle of optimal design. Otherwise, she is not being a cooperative communicator. The fact that she used this description, then, provides evidence that she must have somehow estimated that the mere description “the banana scene” would be enough to
elicit the paroxysms of laughter that now proceed from the other end of the telephone. For instance, she might have assumed that Carla, unlike Bob, was a member of a community that would find this sort of meta-humor especially salient.

The point is that when the mutual knowledge between interlocutors is sufficiently vague, it will always be possible to bring Ann’s description in line with the principle of optimal design, no matter what description she gives. In cases such as these, the vagueness is such that the mutual knowledge between interlocutors can be constructed post hoc to make any description seem like it was designed according to the normative principles of cooperativeness and optimal design. And in real life, more often than not, the mutual knowledge between interlocutors is quite vague. Calculating what interlocutors mutually know a priori is difficult, if not impossible, to achieve with any exactitude, for language users as well as for theorists of language use. A major source of this vagueness is the heuristic of community membership, which is purported to be the primary source of evidence for convention use. What can be inferred to be mutually known on the basis of community membership is much more vague than what could be inferred through physical copresence. How many different communities do individuals mutually belong to? What communities are these, and who are their members? The answers are undefined, because new communities can always be constructed on the fly.6

---

6 Woody Allen himself said, “I wouldn’t want to be a member of any club that would have me as a member.” However, in saying this he is claiming that he is actually a member of the club that wouldn’t have their members as members. Was Woody making a sly allusion to Russell’s paradox?
Therefore, in most everyday situations, the *a priori* mutual knowledge between interlocutors via community membership is inestimable with any degree of certainty.

Due to the inherent inestimability of mutual knowledge based on community membership, one might wonder whether the claim that mutual knowledge is “the essential condition for conventions” (Clark & Carlson, 1981), and other related claims that rely on community membership, can be empirically tested. One possibility is that the theory is simply unfalsifiable, because the mutual knowledge between interlocutors can be constructed *post hoc* such that any behavior can be brought in line with the principle of optimal design. If this is the case, then the theory of mutual knowledge would fall into a class of theories, such as psychoanalytic or literary theory, which provide interesting explanations of phenomena but whose empirical validity is simply irrelevant. This is hardly desirable for a theory of language use.

I believe that the theory is, in fact, testable, but only when two conditions are met: (1) when the mutual knowledge between interlocutors via community membership is explicitly defined and rigorously controlled in the laboratory; and (2) following Keysar (1997), when experiments are adequately designed to address the problem of subsumed explanations. The first condition simply eliminates the possibility of constructing *post hoc* explanations to preserve the normative assumptions of mutual knowledge. If we know precisely the content of mutual knowledge which is relevant for a particular task, then we can decisively measure whether or not that information was used in a particular instance of utterance production or comprehension.
The problem of subsumed explanations, as formulated by Keysar (1997), is that many demonstrations of mutual knowledge have confounded common ground with perceptual or attentional salience. Typically, experiments which claim to show that interlocutors used a piece of mutually known information in production or comprehension do not provide an important control, which is to show that the interlocutor used the information because it was mutually known, and not merely because it was perceptually or attentionally salient. Interestingly, when these controls are implemented, interlocutors’ behavior appears to stray from the standards of mutual knowledge (Keysar, Barr, and Horton, 1998).

The appeal to intentional states is powerful, in that it can explain many aspects of language use, from how communicators formulate definite descriptions, to why their referential descriptions shorten through repeated usage. If we do find evidence that disconfirms that intentional concepts such as mutual knowledge are routinely used in coordination, we are led us back into the quandary which motivated the theory in the first place: how is it that interlocutors coordinate meaning in the face of vast arbitrariness? In the next section, I propose a theory of linguistic coordination that I hope can approach an answer to this question.

1.4 THE THEORY OF DYNAMIC COORDINATION

While the theory of mutual knowledge seeks to explain coordination through the operation of the global mechanism of mutual knowledge, the theory of dynamic coordination explains coordination as a phenomenon that emerges through the operation
of low-level mechanisms. In this section, I introduce the mechanisms of dynamic coordination, with special attention to how these mechanisms are related to the emergence and use of linguistic conventions. Finally, I derive predictions from the theory that the experiments in Chapters Two and Three directly test.

1.4.1 Components of Dynamic Coordination

The theory of dynamic coordination consists of five basic components that guide coordination when interlocutors produce and comprehend utterances. These principles are egocentric anchoring and adjustment, recycled verbal plans, use of prior context, evolution of utterances, and multimodal monitoring. In each case, the proposed mechanisms are directly related to known aspects of cognition and multimodal communication. Thus, the theory does not posit any novel or specialized cognitive processes to explain linguistic coordination. To the degree to which these mechanisms are successful in explaining linguistic coordination, the theory can reduce away some of the higher-level constructs typically invoked in pragmatic theories.

1.4.1.1 Egocentric Anchoring and Adjustment

Formulating and understanding utterances is an uncertain activity. When we converse, we face many decisions about how to describe a referent or how to infer what a communicator means by the use of certain words. Thus, language use can be construed as a kind of decision-making task. One of the great needs of a decision-maker is to reduce uncertainty. In standard theories of language use, mutual knowledge plays this role.
Boaz Keysar and I (Keysar & Barr, in press) propose that language users reduce ambiguity instead by employing an egocentric “anchoring and adjustment” heuristic. As embodied in models of language processing, this heuristic is known as “perspective adjustment” (Horton & Keysar, 1997; Keysar, Barr, and Horton, 1998; Keysar, Barr, Balin, & Paek, 1998). Egocentric anchoring and adjustment is similar to other theories of anchoring and adjustment in decision making (e.g., Tversky & Kahneman, 1974) in that it assumes that language users “anchor” their decision to specific values and then attempt to adjust away from these anchors.

However, unlike other theories of anchoring and adjustment, the value, or meaning, on which interlocutors anchor their judgments is not arbitrary, but egocentric. People formulate and interpret utterances using information that is salient or available to them, and then attempt to use mutual knowledge to adjust to the perspective of their interlocutor. Though they attempt to adjust, the adjustment is typically insufficient. Because of this, both the utterances that communicators produce and the meanings that addressees or overhearers impute to communicators are biased toward the egocentric anchor. The phenomenon has now been established in many domains of language use, including the on-line comprehension of referential descriptions (Keysar, Barr, Balin, & Brauner, in press); the appreciation of sarcasm (Keysar, 1994); the production of referential descriptions (Horton & Keysar, 1997); and the perceived transparency of the meanings of idioms (Keysar & Bly, 1995).

What is the nature of the adjustment process? The theory as currently formulated claims that language users can either spontaneously initiate adjustments or
can adjust when errors are detected from interlocutor feedback. These two possibilities are not mutually exclusive; in fact it seems self-evident that communicators would use feedback signals from their addressees. What remains to be shown is the degree to which they engage in spontaneous, self-initiated adjustment.

Egocentric anchoring and adjustment is a theory that is directly in line with the intrinsically cognitive nature of language use. While egocentric interpretation is often trivially easy, estimating what is mutually known with an interlocutor can be very difficult (as in the above example where Ann attempted to ask Bob whether he’s seen the movie *A Day at the Races*). Anchoring and adjustment appears to be a strategy that is optimal in consideration of the constraints that finite cognitive resources and the time-scale of conversation impose on language use. Though egocentric anchoring and insufficient adjustment may lead to systematic misunderstanding, processes of adjustment can often correct them.

1.4.1.2 *Recycled Verbal Plans*

When Ann first attempts to describe her favorite scene in *Sleeper* to Bob, she has to give a relatively long, elaborate description in order to perform a felicitous act of reference. She could choose an infinite number of descriptions to describe the scene—how is it that she ever finds a description that would meet the criterion of identifiability and uniqueness, and would be consistent with her mutual knowledge with Bob? Descriptions such as these would be most effortful to produce.
In describing the scene the first time, Ann calls into existence a verbal plan that she can store in memory for future uses. Each time Ann reuses it, this “recycled verbal plan” greatly minimizes her effort. Recycled plans are different from precedents in the sense that they are not partner-specific. The theory of mutual knowledge suggests that each time Ann describes the scene to a new addressee, she would have to create an utterance anew, based on her mutual knowledge with that particular interlocutor. However, instead of formulating an entirely new utterance each time, she can simply retrieve directly from memory the description she used with Bob, and then dynamically refashion it should it prove to be unsuccessful. Ann’s final description would have an installment-like character, with the first tone unit consisting of the convention she established with Bob, and the rest of the utterance would be concatenated bit by bit until reference was achieved. This fits with Clark and Wilkes-Gibbs’ (1986) observation that communicators often give descriptions in installments, but not with their assumption that mutual knowledge is criterial for the recycling of previous utterance plans.

1.4.1.3 Use of Prior Context

Choosing a label for a referent is essentially a problem of categorization (Brown, 1958). When naming a dime, we can call it dime, coin, money, extremely rare 1952 dime minted in Philadelphia, object, thing, currency, Fred’s dime, the thing Mary found between the couch cushions, etc. Each label constitutes a different way of categorizing a referent with respect to other objects in the world, and functionally differentiates the referent from an assumed set of background entities. When we call a
dime a dime, we differentiate the object from currency of different value, such as nickels, pennies, and quarters. When we call a dime a coin, we differentiate it from other forms of currency, such as paper currency. The label that a communicator chooses for a referent will depend systematically on the set of objects from which the referent is to be differentiated.

The set of immediate alternatives from which a referent must be differentiated can be termed the local context. Each time that Ann describes the banana scene to a new addressee, she describes it within a new local context, because the overlap in mutual knowledge changes the set of alternatives from which the referent is to distinguished. Accordingly, if she designs her utterance in a way that is sensitive to her mutual knowledge with that addressee, then she should use the copresence heuristics to create slightly different verbal plans in each instance. This would be extremely effortful to do because each time, she would have to start her plans from scratch.

Instead of always tailoring her descriptions to mutual knowledge, Ann could store knowledge about characteristics of the local contexts that she is likely to encounter on any given occasion, based on her past experience. Having used the description “the funny scene with the giant banana” with many different addressees, she might have found that any addressee who has seen Sleeper would remember many funny scenes, but only one that included a banana. In other words, information about the banana was most likely to consummate reference in the greatest number of contexts, while information about its funniness was not. This kind of knowledge can be called prior context. Ann could use this prior context when she faces a new addressee. Importantly, the theory
does not claim that this prior context would have to be mutually known to be included in a description.

When addressees attempt to understand a communicator’s description, they could search for a referent among many different mutually known alternatives, or use information from prior context, even when it is not mutually known. Consider the case of Ann’s old flame, Ernesto, who works at the ticket counter at the Roxy. Ernesto is a connoisseur of comic films, and can enumerate every scene of every movie where a protagonist slipped on a banana peel, from Harpo Marx to Woody Allen. When Ann arrives to the “E” page in her address book, she remembers that Felicia told her last week that Ernesto, with whom Ann has not spoken with in several years, recently started working as a ticket vendor in the Roxy. Therefore, Ann reasons, Ernesto must have seen *Sleeper*, but she could not assume that Ernesto would know that she knew this—that is, she couldn’t assume that this proposition was mutually known. She also fails to consider Ernesto’s expertise on films and does not appreciate that, for Ernesto, “the banana scene” could describe any of several thousand films stored away in his head. When Ernesto interprets Ann’s question, “did you like the banana scene?”, the theory of mutual knowledge would suggest that to find a referent, he should search through referents that are mutually known. In this case, he should consult his fading memories of all of the movies he and Ann saw together during that summer three years ago when he and Ann were dating and he worked in a record store. However, *Sleeper* is currently all the rage, and every night the patrons leaving the Roxy are always raving about “the banana scene”, “that scene with the giant banana peel”, and so on. This makes the association between
the word banana and the film *Sleeper* highly salient for Ernesto. If Ernesto uses this information instead of mutual knowledge, he might understand her definite description right away, simply through prior context—the fact that in 1973 anyone who mentioned a banana scene was most likely talking about Sleeper, and not the one in an early Marx Brothers film. In this way, the reference might succeed, even though it was not mutually known that Ann knew his current occupation—which, according to Clark and Marshall, would be criterial for Ann’s act to be felicitous.

An important consequence of recycled plans and the use of prior context is that descriptions can fail when they violate mutual knowledge. This is not a problem for dynamic coordination, because it posits that interlocutors can apportion effort to adjust their conventional and egocentric plans when needed.

1.4.1.4 *Evolution of Utterances*

When Ann has described the scene to all the friends in her address book from A to Y, and is about to call her new acquaintance Zelda, she has probably learned a lot about the kinds of information would be most likely to guarantee a felicitous act of reference. Certain aspects of the description were likely to evoke positively reinforcing signals from addressees, such as her mentioning of the giant banana, while others failed to spark a single note of recognition, such as “the really funny scene”, these latter descriptions falling short of the presupposition of uniqueness. Aspects of the description that were more likely to successfully consummate reference would become part of recycled verbal plans, thus creating a stronger trace for these aspects in memory, making
them more easily retrieved on the next reference turn. In other words, during repeated reference with different addressees, utterances “evolve” based on the selective mechanism of communicative success and multimodal feedback. This selective process enhances the “fitness” of descriptions in securing reference.

If the only thing Ann knows about her new acquaintance Zelda is that she is a female English speaker who has seen *Sleeper*—that is, if she has just a morsel of local context (mutual knowledge) to base her utterance on—she could maximize the likelihood of successful reference by using descriptions which were successful in the past with other addressees. Importantly, this does not even have to be a deliberate effort on Ann’s part. Without even really trying, she is most likely to include those elements of descriptions that were successful in the past than new ones, because they were used more often and therefore would be more available. Even if these tried-and-true recycled verbal plans fail, Ann and Zelda could still consummate reference when Zelda prompts Ann for more information (e.g., “You mean the scene in Sleeper?”).

This mechanism, intrinsically tied up with effort minimization, explains why it is that the use of the egocentric heuristic and/or prior context can so often succeed. Interlocutors such as Ann and Zelda give stronger weight to those aspects of descriptions that have been successful in consummating reference than they do to those which fail. The selective adaptation of utterances causes them to shorten and become more conventionalized. No special effort would be required for this to occur, because natural cognitive processes of memory retrieval ensure that the verbal plan that is most easily
retrieved is the most likely to be successful with a new partner. In large part, coordination takes care of itself.

Likewise, when addressees successfully retrieve a referent for a description, this strengthens the association between the description and the referent in memory, and makes them more likely to retrieve the same referent when they hear a similar description.

1.4.1.5 Multimodal Feedback Monitoring

Yet how would Ann decide to weight the different elements of her descriptions? One way is through their communicative success. The theory also assumes that she would do this through multimodal feedback from addressees, which would serve as “reinforcement signals” that would positively reinforce some aspects and negatively reinforce others.

Additionally, this feedback can serve to prompt adjustment processes. In contexts where there is an abundance of feedback, as in face-to-face communication, interlocutors can be “lazy”; communicators can wait for addressees to let them know that an egocentric description is inadequate, and addressees can wait for communicators to tell them when their egocentrism is leading them to select the wrong referent. When online feedback is limited or nonexistent as in email, interlocutors might be more careful and spontaneously initiate adjustment to prevent errors.

In brief, multimodal feedback provides a “teacher signal” which interlocutors use to apportion effort to adjustment processes. Communicators use these signals to
adapt their descriptions and in a way that increases their chances of communicative
success while minimizing their effort. Addressees would use feedback to dynamically
adjust their interpretations to better approximate their interlocutor’s perspective.

1.4.2 Implications and Predictions of the Theory

If conventions can be established and used through dynamic coordination,
certain empirically testable predictions can be derived. This section outlines these
predictions and their implications. This sets the scene for the two experiments, which
examine the emergence and use of conventions in language production and
comprehension. The experiments are designed on the analogy of Ann’s use of the
convention ‘the banana scene’ with Bob, Carla, David, Ernesto, and the others in her
address book, from F to Z.

1.4.2.1 Lack of Partner-Specificity

Unlike descriptions formulated under mutual knowledge, descriptions
formulated under dynamic coordination are not necessarily constructed according to the
principle of optimal design. The idea that interlocutors egocentrically anchor and adjust,
recycle plans, evolve maximally “fit” utterances, use prior context in a way that is
insensitive to mutual knowledge, and dynamically apportion effort based on multimodal
feedback suggests that there should be a systematicity in the occasions in which reference
would not be successfully consummated.

We can explain such systematicity as follows. The evolution of utterances
predicts that each time communicators successfully describe a referent, the memory trace
for the verbal plan is strengthened, and the product becomes more codified and easier for the communicator to access in future reference turns. Thus, when communicators face new addressees, the easiest thing for them to do is to recycle the plan she used with the previous addressee. In cases where this description would violate mutual knowledge, or where the addressees’ prior context differed from the communicator’s, this should lead to systematic misunderstanding and the interlocutors would need to apportion a large amount of effort to consummate reference.

1.4.2.2 The “Illusory Transparency of Convention”

If it is true that interlocutors egocentrically anchor and adjust in convention use, then they should believe that the conventions that they use are more transparent than they seem. With a nod to Keysar (1994), we could term this phenomenon the *illusory transparency of convention*.

Conventions seem more transparent than they are because people do not sufficiently adjust their judgments to the perspective of their interlocutors. When communicators produce a conventionalized definite description, they know the referent of that description. Their knowledge of the referent would be difficult for them to ignore when they attempted to assess how their addressee or an overhearer might understand the convention.

From the point of view of comprehension, addressees who experience an illusory transparency of convention would overestimate the conventional meanings of words and underestimate the degree to which the meanings of words reflect local context.
Thus, when Ernesto is discussing a Marx Brothers film with a friend who mentions “the banana scene”, he might be likely to retrieve the scene from *Sleeper* because it is highly associated with this description. He would probably catch his mistake fairly quickly and then retrieve the scene from the mutually known Marx Brothers film.

Keysar (1994) originally discovered the phenomenon of the illusory transparency of intention in an experiment that attempted to examine the use of mutual knowledge in the appreciation of sarcasm. He found that when overhearers had knowledge about a communicator’s attitude that made an utterance sarcastic, they would overestimate the degree of sarcasm that an addressee who did not have this knowledge would perceive. He suggested that this sort of illusory transparency is a routine phenomenon in language use, which predicts systematic misunderstanding in conversation.

However, the theory of dynamic coordination does more than simply document illusory transparency in convention use—it provides an explanation of *why* interlocutors would consider their conventions to be transparent than they actually are. The explanation is based upon interlocutors’ use of prior context and the evolution of utterances.

The model assumes that these processes are relatively inaccessible to language users. Language users do not appreciate the role of communicative success in how they adapt descriptions over time, and are relatively unaware of the degree to which their descriptions are based on local versus prior context. What they are aware of, however, is the ease with which they retrieve the linguistic products that are the result of these
processes. On the first few occasions that communicators describe a referent, they have no choice but to adapt their description to the local cues, possibly including mutual knowledge of the set of objects from which a referent is to be differentiated. However, by the time that they come upon a description that is successful, their attention will have shifted away from the local context that a referent appears in. For instance, if they have described a circle again and again as “the small circle” because the circle always had to be differentiated from a larger circle, they might conventionalize the description “the small circle” and use it even when the local context contains no other circle as a possible referent. Worse, they might use the convention when there is an even smaller circle in the context, leading an addressee who was unaware of the convention to select the wrong referent. An analogous argument would be made for comprehension. The fact that language users would not be aware of this shift in attention from local to prior context would make conventions seem more transparent than they actually are.
CHAPTER TWO

CONVENTION USE IN LANGUAGE PRODUCTION

2.1 INTRODUCTION

The main goal of the experiment described in this chapter is to provide evidence the mechanisms of dynamic coordination in the emergence and use of conventions in language production. These mechanisms are egocentric anchoring, the recycling of verbal plans, the use of prior context, the evolution of utterances, and multimodal feedback monitoring. The experiment is not designed to provide direct evidence for multimodal feedback, though evidence for it can be indirectly inferred.

The chapter begins by summarizing research on conventions in utterance production. It includes a discussion of how communicators use context in generating a message, and then reviews the phenomena of abbreviation and truncation, and the explanations of these phenomena offered by mutual knowledge and dynamic coordination.

2.1.1 The Use of Context in Message Generation

The process of producing a description is generally classified into three phases: conceptualization, formulation, and articulation (Levelt, 1989). In the conceptualization phase, a communicator selects information to be communicated. In the
formulation phase, the communicator transforms this information into a preverbal plan by accessing words, organizing a syntactic structure, and creating a phonetic plan. Finally, the communicator initiates the motor processes required to articulate the utterance.

It is generally accepted that when communicators formulate and articulate utterances, they spontaneously monitor their plans to make sure that the utterances they will or are currently producing are consistent with the ones they planned. While the processes of formulation and articulation operate automatically and with little or no conscious control, conceptualization and monitoring processes are subject to deliberate control, and require attentional resources and working memory. When communicators’ cognitive resources are drained, though they are still able to produce messages, these messages will include more errors because they cannot apportion sufficient effort to conceptualization and monitoring processes.

It is not well understood exactly at what point communicators “assign perspective” to their messages to bring them in line with the perspective of their addressee. The principle of optimal design (Clark & Marshall, 1981) would suggest that communicators consider information such as mutual knowledge and Gricean maxims in the earliest moments of message generation. In other words, the theory of mutual knowledge assumes that the effort that goes into partner-specific tailoring is largely apportioned to the early processes governing the selection of information.

Horton and Keysar (1996) recently tested this idea. They examined how communicators use local context when they describe shapes. In each trial, a target shape appeared in the context of a second shape, or “context shape”. For instance, a circle
might appear in the context of a smaller circle. They manipulated whether or not communicators believed the context shape was shared or privileged. In the shared context condition, they were told to describe the target figure to an addressee who could also see the context figure; in this case, they could assume that it was mutually known, and should describe the target as the “large circle”. In the privileged context condition, they were told that the addressee would not be able to see it; therefore, they should call the target shape the “circle”.

What they found was that communicators were more likely to use local context in the shared condition than in the privileged condition. In other words, communicators’ descriptions in the shared condition were sensitive to mutual knowledge, as the principle of optimal design would suggest. They termed this model Initial Design. However, it was also possible that they formulated their utterances egocentrically, taking into account the context figure, but through internal monitoring were able to filter out this information before articulating their verbal plan. This was called the Perspective Adjustment model.

To test these two models they placed other communicators under a time deadline by which they had to start producing their descriptions. They made the reasonable assumption that if mutual knowledge is routinely considered during conceptualization, then it should be less compromised by this temporal deadline than if it is used to monitor the products of conceptualization. If, instead, it is used by the more deliberate and controlled processes of utterance monitoring, then it should be very heavily affected by the imposition of the deadline.
Horton and Keysar found that under the deadline, participants were just as likely to use context when it was privileged as when it was shared. In other words, participants in the privileged condition were just as likely to call the target “the small circle” as the participants in the shared condition, even though the context figure was not mutually known. These results supported Perspective Adjustment over Initial Design. They concluded that communicators generate egocentric messages, and adjust these messages through internal monitoring when sufficient cognitive resources are available.

2.1.2 Repeated Messages and the Evolution of Conventions

While Horton and Keysar examined the use of context in the formulation of individual definite descriptions, the current experiment contributes to this area of inquiry by examining how the use of context changes over repeated acts of reference, as communicators conventionalize descriptions. This section begins by discussing the conventionalization phenomena that are typically observed in repeated reference, then reviews studies which have sought to explain these phenomena.

2.1.2.1 Abbreviation

In a set of seminal papers, Krauss and Weinheimer (1964, 1966) examined how communicators’ descriptions changed over the course of various reference turns. In these studies, they established the general referential communication paradigm in which a communicator describes a referent to an addressee. One of their findings was that over the course of repeated interaction, descriptions tended to get shorter (Krauss and Weinheimer, 1964). In a follow-up study, Krauss and Weinheimer (1966) had
communicators describe referents to addressees under conditions of variable feedback.

Half of the addressees were given concurrent feedback regarding their performance over 50% or 100% of the trials, and the other half received no feedback. The descriptions of communicators in the concurrent feedback condition became much shorter over time than those who were not given feedback by the addressee.

The fact that addressee feedback affects description length can be accounted for by the collaborative model (Clark & Schaeffer, 1989; Clark & Wilkes-Gibbs, 1986; Clark & Brennan, 1991; Brennan & Clark, 1996). This model suggests descriptions shorten through the establishment of conventions, governed by the accumulation of mutual knowledge. In each reference turn, communicators present descriptions and elaborate upon them as necessary. Addressees, in turn, provide the communicator with signals that they accept or reject portions of the descriptions. In this way, interlocutors minimize collaborative effort, which is “the work that both speakers and addressees do from the initiation of the referential process to its completion” (Clark & Wilkes-Gibbs, 1986). That this effort is collaborative suggests that communicators are not simply interested in minimizing their own effort, but also that of their addressees. In Clark and Wilkes-Gibbs (1986), directors described abstract, human-like shapes to addressees. They found that the description lengths showed an exponential decay from one reference turn to the next, which appeared to support the claims of the collaborative model. The interpretation was that over time, the interlocutors arrived at mutually accepted descriptions of referents that minimized collaborative effort.
Clark and Brennan (1991) later refined this model, introducing the idea of *grounding*. In grounding, when communicators describe a referent, they seek positive evidence that their descriptions have been understood. As in the collaborative model, they first *present* a description to an addressee, who in turn, will either *accept* or *reject* it. The portions that are accepted by the addressee will be recycled in future descriptions, and those that are rejected will fall by the wayside. In this way, communicators and their interlocutors accumulate common ground in an orderly manner regarding the kinds of descriptions that are acceptable for different referents. Because each reference turn is an instance of grounding, it provides an opportunity for further negotiation about how a referent is to be described. The process of collaborative grounding makes interlocutors members of a minimal community of “two”, defined by their mutual knowledge of the conventions. As descriptions become more optimal with respect to mutual knowledge, the descriptions minimize collaborative effort by requiring less use of local context.

A prediction can be derived from the idea that abbreviation is a process of the accumulation of mutual knowledge, or mutual acceptance of aspects of a description. If mutual knowledge is wholly responsible for abbreviation, then it follows that when communicators face new addressees who are not members of this community, the descriptions that they produce should be as lengthy as the ones they gave to the previous addressee during the first reference turn. This is because none of the elements of the previous descriptions have been mutually accepted by the addressee. The experiment in this chapter tests this prediction.
2.1.2.2 Codification

Over time, as partners repeatedly refer to the same referent, not only do their descriptions get shorter, but they also become more ‘codified’—they are less diverse in the lexical elements that they include. Additionally, when interlocutors repeatedly shift roles from addressee to communicator, they can be observed to converge on how they describe certain referents. This phenomenon is known as lexical entrainment. The collaborative model would explain lexical entrainment in the same way it explains abbreviation—by the accumulation of mutual knowledge through the collaborative process of grounding.

Garrod and colleagues have challenged the idea that the accumulation of mutual knowledge is wholly responsible for the process of conventionalization. Garrod and Anderson (1987) argued that in addition to mutually accepted global conventions, “language processing in dialogue may be governed by local principles of interaction...” (p. 181). They state that it seems likely that language users, in addition to mutual knowledge, “have developed processes which take advantage of the inherently interactive and collaborative nature of dialogue to reduce the complexity of inferences required for truly coordinated understanding in everyday conversation.” (p. 185). In many ways, their position on coordination is consistent with the goals of the theory of dynamic coordination.

In Garrod and Anderson’s experiments, participants communicated to one another about positions in a maze. These positions could be conceptualized according to different mental models. The mental models specify how referential terms, such as row
and *line*, or spatial terms such as *over* and *above*, would be mapped onto referents. These mental models were classified into four types. Garrod and Anderson found that over time, pairs would tend to “entrain” upon a certain mental model of the maze. Interestingly, interlocutors appeared to establish these “agreements” tacitly—that is, without explicit negotiation. Even when a particular scheme was explicitly negotiated, it would often be abandoned soon afterward.

Garrod and Anderson concluded that semantic coordination is partially driven by what they termed the *input-output coordination principle*, which states: “formulate your output (i.e., utterances) according to the same principles of interpretation (i.e., model and semantic rules) as those needed to interpret the most recent relevant utterance.” (p. 207). This coordination principle would promote consistency between the mental models of interlocutors.

Brennan and Clark (1996) argued that the kind of lexical entrainment observed in Garrod and Anderson’s studies reflect the establishment of conceptual *pacts* between interlocutors rather than the operation of the input-output coordination principle. In this paper, Brennan and Clark made strong claims about the role of mutual knowledge in conventionalization. They noted that strict application of the I/O principle would lead to a rigidity of description that is typically not seen in conversation, not even among the dyads observed by Garrod and Anderson. Instead, they argued that the interlocutors establish “conceptual pacts”, which they define as mutually accepted agreements about how a referent is to be conceptualized. These pacts determine lexical choice in utterance production. The defining characteristic of conceptual pacts is their “partner-specificity”: 
that is, communicators only use conceptual pacts when speaking to interlocutors with whom these pacts are mutually known. In an experiment Brennan and Clark showed that when communicators face a new addressee, they abandon the conceptual pacts they had established with previous addressees. They claim that this would not be predicted by the input-output coordination principle.

Brennan and Clark interpreted their participants’ abandonment of conceptual pacts as evidence for their partner-specificity. However, they acknowledged that these pacts were not uniformly nor immediately abandoned by communicators when they were confronted with a new partner. Rather, communicators gradually adjusted their descriptions to form new “conceptual pacts”. Though they explained this through the collaborative model, it is not clear how much of the adjustment could have been explained by egocentric anchoring, with adjustment through multimodal feedback from the new interlocutor. Therefore, Brennan and Clark’s results do not allow us to conclude that their participants were truly forming partner-specific conceptual pacts, because communicators simply could have been coordinating through egocentric anchoring and dynamically apportioning effort as needed to correct or prevent errors.

Interestingly, the collaborative model’s emphasis on grounding appears to be at odds with the notion of optimal design. In a world in which communicators optimally design utterances for their addressees, and addressees exclusively use common ground as the context in which they comprehend utterances, why should language users require positive evidence of understanding? The theory of mutual knowledge suggests that this is because the common ground between interlocutors might be difficult for
communicators to estimate with accuracy, or that they might have mistaken beliefs about what is commonly known. However, instead of trying to estimate these vague and indefinite quantities, it might be simpler for communicators just to fall back on egocentric plans and dynamically update them as needed. In this manner, they would converge toward an optimal apportionment of effort to different components of production.

2.1.3 Dynamic Coordination in Convention Use

Though conventionalization could occur through the application of the principle of input/output coordination or by establishing conceptual pacts, the question arises as to the need to postulate specialized mechanisms for that which might be explained more simply through the application of general cognitive mechanisms. The theory of dynamic coordination attempts to explain the process of conventionalization through egocentric anchoring, effort minimization and apportionment, and dynamic weighting of information based on multimodal feedback.

2.1.3.1 Abbreviation: Recycled Verbal Plans

As Krauss and Weinheimer (1966) point out, the truncation process operates most effectively when communicators are given concurrent feedback from the addressee. This fits with both the collaborative model and dynamic coordination. For the collaborative model, descriptions shorten because collaboration enables certain aspects of a description to become mutually accepted.

On the other hand, dynamic coordination predicts that descriptions shorten because communicators are attempting to minimize effort by recycling those aspects of a
description that are successful in establishing reference and jettisoning those that fail. While mutual knowledge theory assumes that it is the mutual knowledge between interlocutors that changes over the course of reference turns, the theory of dynamic coordination assumes that it is, in fact, actual verbal plans for descriptions which are stored, recycled, adapted, and refashioned over the course of reference turns. Each refashioning dynamically weights information according to its communicative success as gauged through the addressee’s multimodal feedback. In other words, the addressee provides the selective mechanism that determines how utterances evolve. This has the result that utterances which are more likely to succeed become easier to produce. Because these plans retain the successful elements and lay aside the bad, communicators can reach optimal coordination while minimizing effort. Over time, a description will evolve such that it appears ‘tailored’ to the mutual knowledge between interlocutors, because information which is mutually known is more likely to engender successful acts of reference. Yet even though it appears tailored for a particular addressee, this is through evolution via selective reinforcement, but not of initial design.

2.1.3.2 Codification: The Use of Prior Context

A related way that dynamic coordination theory predicts that conventionalization occurs is through communicators’ use of prior context. When communicators encounter a referent in a new context, they have to figure out what information to include in their descriptions in order to uniquely identify the referent within that context. Prior context is effort minimizing, because it limits the amount of
new contextual information that communicators must consider in a given reference turn. Over time, communicators would differentially weight aspects of a description according to the number of contexts in which it resulted in a successful act of reference. By tracking the predictive value of this information, communicators could maximize their chances of meeting the presupposition of uniqueness. The use of this information would also cause communicators’ descriptions to become more and more codified over time, as the variability in their descriptions decreases.

To provide an example, the experiment described below uses geometric shapes such as circles, squares, etc. as referents in a referential communication task. One kind of shape—for instance, circles—appears with a context shape of a different size, such that describing the size of a circle would result in unique reference more often than describing its shade. A second shape, such as a square, is presented in contexts where there is always another square of the same size, but of a different shade. In this latter case, describing the shade of a square would result in unique reference while describing its size would not. If communicators track and make use of this distributional information, then they should show a tendency to conventionalize the related scalar terms, using them even when there is no other circle in the display.

2.2 METHODS

2.2.1 Task

The experiment used a modified referential communication task in which one participant described geometric shapes to another. The two participants sat side-by-side,
facing a computer screen. The participants were randomly assigned to one of two roles: “director” and “addressee”. The participants maintained the same roles throughout the experiment. During the experiment, directors’ eye position was tracked to examine the use of local context in utterance planning and monitoring.

The experiment was divided into a series of rounds. In each round, a set of shapes would appear on the computer screen in the spaces of a 4 x 4 grid. One of these shapes was designated as the “target shape”. The director’s task was to describe the target shape such that the addressee would be able to identify it based on the description.

The task required that only the director know which of the shapes was designated as target. For this purpose, the spaces in the grid were numbered from 1 to 16. Before each round, the numbers appeared in the spaces of the grid and the director heard one of them spoken through an earphone. The shape that appeared in the corresponding square was the target. The earphone ensured that only the director would know beforehand which shape was the target.

After the shapes appeared, the director was free to initiate the description when ready. A restriction was that the director would have to identify the object without pointing to it or referring to its location (e.g., without saying, “the shape in box number 13” or, “the one in the lower left corner”). Thus, the director would have to name or somehow describe the target so that the addressee would be able to identify it uniquely.

Certain experimental items were designed so that when directors formulated descriptions, they would need to consider the presence of contextual objects of the same shape as the target. For instance, if the target shape was a circle, there may be another
circle of a different shade or size, such that it would be insufficient to merely call it “the circle”. Instead, the director would have to mention whatever aspect of the target differentiated it from the other circles in the display (e.g., “large” or “light” circle).

Based on the directors’ descriptions, addressees chose the shape they believed to be the target. Addressees indicated their choices by clicking on the shape with the computer mouse. When addressees chose the correct shape, it would momentarily turn green. When they chose the wrong shape, it would turn red. In this way, the dyad received feedback on their performance.

2.2.2 Design

This experiment investigates two main questions regarding language production, one regarding the evolution of conventions and the second regarding the partner-specificity of these conventions. On the one hand, it examines how directors’ descriptions change over time—specifically, how they become conventionalized—and establishes the mechanism responsible for these changes. On the other, it probes the partner-specificity of these conventions. Accordingly, there were two independent variables: reference turn, or the number of times that a director has had to describe a shape within a certain kind of context, and community membership, which refers to whether the addressee is a member of the community that knows the conventions.
Figure 2.1. Design of Experiment 1.

The experiment was divided into two phases. The phases are referred to as the acquisition and transfer phases. The two phases of the experiment are schematized in Figure 2.1. The acquisition phase provided the director with the opportunity to conventionalize descriptions over the course of various reference turns. In the acquisition phase, the director described targets to the experimenter, who took the role of addressee.

The acquisition phase itself was divided up into eight smaller “blocks” of trials (the small boxes in the top left of Figure 2.1). Each block consisted of four probe items, ten training items, and three filler items (described below). The four probe items were
presented in a random order at the beginning of each block, followed by the training and filler items, mixed together and randomized.

The transfer phase tested the partner-specificity of the conventions established during the acquisition phase. During this phase, the director described the shapes to an addressee who was absent during the acquisition phase. Therefore, in the transfer phase, directors could not infer mutual knowledge of the conventions through community membership. Instead, directors could only infer mutual knowledge of the physically co-present information contained within each visual array.

2.2.3 Procedure

When participants arrived at the laboratory, they were assigned the role of either “director” or “addressee”, and took their respective seats. The director was seated approximately 0.5 m away from the computer monitor.

The experimenter then explained the task. He told them that they would be playing a communication game in which the director would describe shapes to two different addressees, and that the addressee would select shapes based on these descriptions. He then explained the earphone setup and advised the directors that they were not permitted to identify shapes by pointing to them or by referring to their locations.
Next, the participants were given five practice trials in order to familiarize them with the procedure. The target shapes used in the practice trials did not include any of the critical targets used in the transfer phase—only filler shapes were used as targets.\footnote{One of the filler shapes served as target in two of the practice trials. This shape was a square rotated 45 degrees such that it appeared as a diamond. I naively assumed that all participants would refer to this shape as the diamond; however, a few participants called it “the square”. Nonetheless, directors nearly unanimously settled upon some alternative description for this shape during the acquisition phase.}

After the practice trials, the experimenter introduced the eyetracking equipment. The operation of the equipment was described. However, to avoid drawing undue attention to eye-movements, participants were not informed that the eye camera tracked the position of their eye. Instead, they were informed that it “records features of the eye such as the diameter of the pupil”. The participants were given the opportunity to ask any questions about the experiment. Once it was clear that they understood the task, they were given consent forms to sign.

Next, the experimenter mounted the eyetracking equipment on the director’s head and calibrated the equipment.

2.2.3.1 Acquisition Phase

Over the course of the experiment, directors described shapes to two different addressees—the experimenter and another participant. In the acquisition phase, directors described shapes to the experimenter. In the transfer phase, directors described shapes to an addressee who was not present during the acquisition phase, and therefore could not know the conventions established between the director and the experimenter. So that the
addressee would not hear these conventions, before the acquisition phase began, the addressee was led down the hall into a waiting room.

In the acquisition phase, the director described the target shapes to the experimenter, who in turn, clicked on them using a computer mouse. Between trials, the trial and subject numbers flashed in the upper left hand corner of the screen to allow the experimenter to coordinate the video with the digital data during analysis.

After completing all of the blocks of the acquisition phase, a message box appeared on the computer screen which invited the director to relax and take a two-minute break while the computer loaded the second half of the experiment. The message box also advised the director to refrain from discussing the experiment during this period. The purpose of the break was to give the experimenter time to go down the hall and retrieve the other participant, who would serve as addressee for the transfer phase.

The acquisition phase had eight blocks of 17 trials each, for a total of 136 trials, requiring approximately 15 minutes to complete.

2.2.3.2 Transfer Phase

After the two-minute break, the transfer phase began. The purpose of this phase was to examine whether the conventions established in the first phase were “partner-specific”, or whether they simply represented effort-minimizing strategies. Directors would not be able to infer that these conventions were mutually known with the new addressees, and should rely on physically co-present information in formulating descriptions for them.
After completing the procedure, the participants were given the opportunity to guess the purpose of the experiment. Through several open-ended questions, the experimenter sought to ascertain whether participants were explicitly aware of the shape-contrast pairings from the first part of the experiment, and then thoroughly debriefed them. Most directors reported not having noticed any systematicity in the size-shade pairings of the scalar trials.

The entire procedure lasted approximately 35 - 40 minutes.8

2.2.4 Materials

2.2.4.1 Acquisition Phase

The stimuli used in the referential communication task were grayscale images of geometric shapes. The shapes were drawn on a computer screen on a black background. They were drawn with a white outline and filled with a certain percentage of white pixels, yielding a uniform shade of gray. The shapes varied along two continuous dimensions: size and shade. The size of shapes was computed relative to the size of one unit in the grid. In this experiment, the size of one grid unit on the monitor was approximately 4.93 cm by 4.93 cm. The shapes varied in size from approximately 22.2% (1.09 cm) to 77.8% (3.84 cm) of a single grid unit. The shade of shapes, based on the percentage of white pixels on a black background, ranged from 23.5% to 78.4%.

---

8 After the transfer phase and before debriefing, typicality judgments were also collected, which are not reported in this document.
The shapes fell into three different classes depending on their purpose in the experiment: nominal, scalar, and filler shapes. The first class, nominal shapes, consisted of two shapes that were difficult to describe, due to their novelty and lack of conventional names (shown in Figure 2.2). They varied in size and shade, though whenever they appeared as targets they appeared in the absence of other shapes of the same category. The only stable feature of these shapes across context was shape—size and shade varied randomly. The purpose of these shapes was to investigate how descriptions along a discrete dimension (shape) are conventionalized over the course of many reference turns.

![Nominal Shapes](image)

**Figure 2.2. Nominal Shapes.**

The second class of shapes used in the experiment was included to examine how descriptions along continuous dimensions (size and shade) are conventionalized over time. Unlike the abstract shapes, the two shapes in this category have conventional, everyday names: “circle” and “square”. However, when these shapes appeared as targets, they typically appeared in the presence of another shape of the same category that varied along one of the two dimensions. For instance, if a circle appeared as a target, it would
typically appear with another circle that contrasted either in size or shade (but not both). Thus, directors could not simply name the target shapes (‘circle”), but had to refer to them contrastively, using a scalar adjective (e.g., “small circle”). For this reason, these shapes are called scalar shapes.

**Figure 2.3. Scalar Target Shapes.**

The scalar shapes used as targets in the acquisition phase were of two possible sizes and two possible shades, for a total of four exemplars from each category of shape. The exemplars are displayed within “size-shade” space in Figure 2.3. The horizontal axis represents the size dimension, while the vertical axis represents shade. Exemplars were staggered within the space such that a relatively “large” square would be of a different absolute size than a relatively “large” circle. Likewise, a “light” circle is of a different
shade than a “light” square. This was to encourage the relative over absolute use of the scalar terms large / small, and light / dark.

Each of the scalar exemplars appeared 57 times during the experiment during the acquisition phase, and from 1 – 5 times in the transfer phase.

In addition to squares and circles, there were five kinds of filler shapes. These shapes are shown in Figure 2.4. The size and shade of these shapes was allowed to freely vary within the following ranges: size: 1.78 cm (36.1% of one grid unit) to 3.84 cm (77.8%); shade: 37.3% white pixels to 78.4%.

![Figure 2.4. Filler Target Shapes](image)

In the acquisition phase, there were two main sorts of items (visual displays): training and probe items. Training items, which were used only for scalar shapes, gave directors the opportunity to gradually acquire information about the contexts in which scalar shapes typically appeared. Examples of training items are shown in Figure 2.5. In the training items, a target scalar shape was always paired with a shape from the same category—henceforth contrast shape—which contrasted with the target in either size or shade.
Consider the shapes in Figure 2.5. In the training items, target circles always appeared with a contrast circle of a different size. Thus, when referring to these targets, directors would always have to use a scalar term referring to the relative size of the shapes (e.g., “small”, “little”, “large”, or “big”). The constant pairing of circles in a size relation was intended to focus the director’s attention on the size dimension when describing circles. Target squares, on the other hand, always appeared with a contrast square of a different shade.⁹ This would require directors to use terms such as “light” or “dark” when referring to a square.

![Figure 2.5. Examples of Scalar Training Items.](image)

Each time a particular kind of target shape appeared, it always appeared with the same kind of contrast shape. Thus, a target circle would always appear with a circle

---

⁹ These pairings were reversed in a second version of the items.
of a different size but the same shade, while a target square would always appear with a square of a different shade but the same size. Over time, directors could use their knowledge of the kinds of contexts that the scalar shapes were likely to appear in to establish conventions, such that a certain size of circle would be the “small circle” and a certain shade of square would be the “light square”.

There were a total of 80 training items in the experiment, 10 for each of the eight scalar shapes. In each item, a random number of filler shapes were added (from 7 to 10) to make it effortful for the director to exhaustively check all of the shapes in the context.

While the training items gave directors the opportunity to accumulate knowledge about prior context, the probe items revealed the degree to which directors used this knowledge in the form of conventions. The target shapes for the probe items were drawn from the set of two nominal shapes and the set of eight scalar shapes. Unlike the training items, the probe items did not include other shapes that matched the target shape; i.e., there were no contrast shapes. Instead, the grid was filled with objects of a different shape than the target. The sizes and shades of these filler shapes were determined randomly.

The probe items with scalar shapes as targets are critical for testing the main hypotheses of this experiment, because they revealed the degree to which directors, when they formulated descriptions, used conventions they had established with the previous addressee, instead of properties of the visual array (local context). Following Lewis (1968), conventions are operationalized as alternatives that, by virtue of mutually
established precedence, override more salient alternatives. In a given reference turn, the director has available two sources of mutual knowledge that could guide the preverbal plan. On the one hand, the director could appeal to information that is physically copresent with the addressee—the identity of shapes in the grid. On the other hand, the director could use the mutually known prior context—the conventions that have been established over the course of many reference turns. Thus, when a circle is presented alone, the most salient alternative might be to refer to it simply as “the circle”. However, when that same circle has been referred to repeatedly as the “small circle”, the director could instead opt for this description even though there are no other circles in the display. Moreover, the director would minimize effort by doing so, because it would be unnecessary to scan the display for contrast objects. Such a description would not be redundant for addressees who could infer mutual knowledge of the convention through community membership.

There were a total of 32 probe items in the experiment. Sixteen of these items contained one of the two nominal shapes as targets. The other 16 items contained one of the eight scalar shapes as targets. The probe and training items were organized into eight blocks (see Figure 2.1). Each block consisted of four probes followed by ten training items. Two probe items presented the nominal shapes as targets, and the other two presented two of the eight scalar shapes as targets. Over the course of eight blocks, each scalar shape was seen as target twice, with the order in which they appeared randomized across blocks. Each nominal shape was seen eight times, once at the beginning of each block. The order of the four probe trials within a single block was also randomized.
In each block, the four probe items were followed by the training items. Three “filler” items were mixed in with the training items. In the filler items, one of the five filler shapes appeared as target, with no relevant contrast shape. The filler items performed two functions in the experiment. First, they helped obscure any explicit awareness of the shape-contrast pairings. Secondly, they made it difficult for the director and addressee to adopt the strategy of only paying attention to the scalar and nominal shapes in the grid, because these shapes appeared as contextual shapes in some trials.

2.2.4.2 Transfer Phase

The transfer phase was designed to test the “partner-specific” aspect of conventions. Directors described target objects to new addressees. Because these addressees were absent during the acquisition phase, directors could not infer mutual knowledge of the conventions that they established with the experimenter. According to mutual knowledge theory, directors should seek to optimally design their utterances with respect to their mutual knowledge with the addressee. Therefore, they should consider only physically co-present information when formulating their descriptions. Otherwise, they run the risk of violating the Maxim of Quantity and make the addressee’s task of identifying the target more effortful than necessary. For instance, when a speaker calls the only circle in the display “the small circle”, this description carries the presumption that there is another circle in the display which is larger. Because of this, the addressee would not be able to terminate referential search immediately upon finding the lone circle, because the description presupposed that there were at least two in the display.
When directors faced new addressees, consider how they should alter their descriptions of the nominal shapes to be consistent with mutual knowledge. Let us assume that when the director first saw the shape on the left-hand side of Figure 2.2, she gave a highly elaborate description of it, such as “the weird looking dark gray thing with kind of a protrusion to the right side... it looks like three rectangles kind of stacked up unevenly to form a tetris-like shape”. Let us assume further that over the course of the acquisition phase, she had shortened this description to merely “the tetris shape”. Under the common ground hypothesis, when she first describes the target shape to a new addressee, her description should be as elaborate as it was in the first reference turn. On the other hand, if she is recycling verbal plans, she may describe the shape as, “the tetris shape”, and then adjust it as needed to secure reference. Dynamic coordination theory predicts that her description should be shorter than the one she gave during the first reference turn in the acquisition phase, but it should also be longer than her last description with the previous addressee, due to adjustment.

To test between these alternatives, six transfer items used the nominal shapes as targets, three for each of the two nominal shapes.

Like the nominal trials, the scalar trials also investigated the degree to which directors altered their descriptions for new addressees. During the acquisition phase, it is hypothesized that there would be an increasing tendency for directors to refer to a particular size of circle as the “small” or “large” circle, even in the absence of a contrast shape. As in Lewis’ (1968) definition of conventions, the precedent overrides a more salient description based on local information. Once directors felt relatively secure that
they had seen exemplars of the entire set of objects, they could start using conventions, because the conventions would obviate the need to fully check the display for contrast objects. Using conventions would also minimize effort for those addressees who are also aware of the conventions, because by the time they heard the scalar “small” or “large”, they would be able to guess which shape was being referred to. However, addressees who were unaware of the conventions would try to make the scalar term relevant by looking for a shape that had a contrast pair in the display.

To understand why this is so, consider the finding from Tanenhaus and colleagues that addressees use visual context along with linguistic information to incrementally constrain the domain of reference (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1996). Data from eyetracking experiments suggested that when people comprehend referential descriptions, linguistic information is immediately integrated with visual context to limit the set of possible alternatives. Tanenhaus and colleagues observed that addressees were able to locate referents even before hearing the full noun phrase. Imagine how an addressee, who is naive to the conventions, would understand the phrase “the small circle” in the context of Figure 2.6. By the time the addressee heard the scalar adjective “small”, he or she could hypothesize that the director must be referring to the smaller of the two squares in the figure, because the use of “small” suggests the presence of a contrast. This would be indicated by an eye movement to this target object. Yet once the addressee heard the word “circle”, he or she would have to revise this hypothesis.
The scalar transfer items were designed such that directors would produce different descriptions if they used the conventions established in the acquisition phase instead of the physically co-present information in the visual array. At one extreme, directors could be completely egocentric and continue using the conventions they established with the experimenter. At the other, they could tailor their descriptions to mutual knowledge they share with the new addressee. Given that the addressee does not know the conventions, directors would need to make use of the physically co-present information in the visual display. Egocentric anchoring with spontaneous self-initiated monitoring predicts that directors’ use of conventions after switching partners should drop somewhat as they attempt to adjust to the new addressees, but not as low as the baseline level because this adjustment will be insufficient and biased toward the egocentric anchor.
Ten of the transfer phase items included scalar shapes as targets. To subvert directors’ ability to predict the kinds of contrasts targets would appear in, three different kinds of items were created. These three categories were defined by the presence versus absence and/or nature of contrast objects: four unique shape items; four unconventional contrast items; and two extreme contrast items. Examples of these three different kinds of item are shown in Figure 2.7.

(a) unique shapes | (b) unconventional contrast | (c) extreme contrast

**Figure 2.7. Scalar Transfer Items.**

In the unique shape items, a scalar target was presented in the absence of any other contrast shape. The question of interest was whether or not the director would refer to the target shape with the unadorned name (e.g., “the circle”) or use the conventionalized scalar (“the small circle”). The use of the convention would not be optimally designed for the mutual knowledge with the addressee. For instance, if the director referred to the target square in Figure 2.7a as “the light square”, then an addressee might consider the light circle as a referent momentarily before hearing the full phrase.
In the *unconventional contrast* items, the scalar targets appeared accompanied by contrast shapes. However, unlike the training items, the contrast shape differed from the target along the unconventionalized dimension. Let us assume that the target in Figure 2.7b had been conventionally referred to as “the small circle”. Though this description would have been adequate in the acquisition phase, it would be inadequate here, because there is a contrast circle present that differs in shade. Thus, reliance on the convention alone would lead to an uninformative description. If directors are completely egocentric in formulating their descriptions, then they should make this kind of mistake. If, however, they fully take into account common ground and base their descriptions purely on physically co-present information, then they should call the target simply “the light circle”. Finally, anchoring and adjustment would predict that they would produce a mixed description, such as “the small light circle”.

In the eight unique and unconventional contrast items, each of the eight exemplars appeared once as target.

The final kind of transfer item was the *extreme contrast* item. These items differed from the other two kinds of items because use of the convention in these cases would not only violate a normative standard of informativeness, but if it went uncorrected, would actually lead the addressee to select the wrong shape as the target. Consider Figure 2.7c as an example. The target in this item is a shape that would have been conventionally referred to as “the small circle” in the acquisition phase. However, calling it “the small circle” in this case would possibly lead to an error, because the addressee would interpret “small” as selecting the smaller of the two circles in the
display. Thus, if directors are purely egocentric, then there should be many errors on these items. On the other hand, if they tailor their descriptions exclusively to physically co-present information, then they should say, “the large circle”. However, directors may not be able to sufficiently adjust away from the conventions established in the acquisition phase, as the anchoring and adjustment hypothesis would predict. In this case, we should see extensive use of comparative forms, such as calling the target “the larger circle”. This is because “larger” suggests that although the target is the larger circle in the display, it is still not as large as the “largest” circle in the entire set of shapes. In this way, the prior knowledge that the director has about the set of objects would still figure into the description.

Two concerns arose due to the inclusion of the extreme contrast items. The first concern was that they would cause errors, thereby inducing directors to increase their vigilance of local information, which would carry over to the other transfer items. For this reason, only two extreme trials were included in the transfer phase; one in which a smaller circle that the conventionalized one appeared as a contrast shape, the other in which a darker square than normal appeared. The second concern was that any mistakes would not be due to the lack of attention to local information, but simply to the fact that directors had not previously encountered a circle as small or a square as dark before. To circumvent this problem, the two extreme contrast shapes appeared as contextual shapes
in eight filler items during the acquisition phase, four times each. This reduces the likelihood that any effects would be due purely to the novelty of these shapes. \textsuperscript{10}

\subsection*{2.2.5 Apparatus}

In this experiment, stimulus items were presented on a computer screen and two channels of data were considered: eye movements and verbal utterances. Presentation of the visual stimulus and initiation of recording of eyetracking and audio data were synchronized by means of software specially designed for this purpose in Microsoft Visual C++.

Eyetracking data were recorded in two formats, digital and video. The real-valued coordinates of the eye were recorded in digital format and stored on the hard disk of the eyetracking computer, while the video image of the display was recorded to Hi-8 video.

The audio data for each trial were recorded digitally. The files were recorded in 16-bit, mono WAVE format, at a sampling rate of 11.025 kHz. After recording the data for a given trial, voice onset time (VOT) was computed by using a “moving window” algorithm, described fully in section 2.3.1.2.

\textsuperscript{10} These items were removed from the final analysis because they turned out to be qualitatively different from the other items, prohibiting a reasonable comparison with the baseline values from the acquisition phase. They are described here for the sake of completeness. The results for these items supported anchoring and adjustment: 47\% of directors used the comparative to describe what was previously the “small circle”.
2.2.6 Participants

Sixteen participants, eight females and eight males from the University of Chicago community, participated as directors in the study. The addressees were ten females and six males, also from the University of Chicago. Six of the female directors participated with female addressees during the transfer phase, the other two with male addressees. Four of the eight males described shapes to female addressees during transfer, and four described shapes to male addressees.

Participants in the experiment were recruited by internet postings and flyers posted around campus. They were paid for their participation in the experiment. Of the sixteen pairs, three pairs mentioned that they were either acquaintances or friends. The rest were strangers. Directors ranged from 20 to 30 years of age, though two of the female directors were over 40 (41 and 49 years).

2.2.7 Analysis

2.2.7.1 Dependent Measures

This section describes the dependent measures used to test hypotheses. Though data were recorded for all items, including fillers and training items, the dependent measures were computed only for probe and transfer items. Different measures were used for nominal and scalar shapes, though some measures were used for both. There were a total of five different measures considered in this experiment, listed in Table 2.1 for convenience.
Table 2.1. Dependent Measures, Experiment One.

For the nominal trials, three measures were considered: number of words, lexical diversity, and voice onset time.

**Description length.** This measure is simply the number of words in the director's description, excluding hedges such as "um" and "uh". The raw number of words used in a description is an index of the amount of information that the director believes the addressee needs to successfully identify the target object.

**Lexical diversity.** While the length of a description gives some indication of conventionalization, it is possible that directors choose different, albeit shorter, descriptions during each reference turn. The second measure, lexical diversity, reflects the degree to which descriptions recycle lexical units from one reference turn to the next. In essence, it reflects the “codification” of descriptions over time. Lexical diversity is simply the proportion of words in a given description that did not appear in the previous description. Imagine, for instance, that during block 5, a director refers to an object as “the large tetris thing” after having referred to it in block 4 as “the large tetris shape with the point sticking out”. Three of the four words in the block 5 description occurred in the immediately previous turn (“the”, “large” and “tetris”), and only one word—“thing”—
was not used in the previous turn, such that the lexical diversity of this description would be 1/4, or 0.25. Descriptions from block 1 were automatically assigned a value of 1.0 (maximum possible diversity).

This measure is less useful than number of words in testing hypotheses, because of the lack of a true baseline. The lexical diversity is automatically set to 1.0 at block one, and it would be unlikely that a description would ever have zero overlap with the previous description. On the other hand, this measure is useful for exploring how quickly directors codify their descriptions and for verifying that they are not merely choosing different, albeit shorter, descriptions each time.

**Voice onset time (VOT).** Voice onset time, or VOT, was computed for both nominal and scalar trials. It is a measure of how long it takes the director to initiate the referring noun phrase, measuring from the moment that the visual stimulus appears. This measure relates to the director’s effort in conceptualizing a message, formulating a preverbal plan, and internally monitoring the plan for adequacy. The use of this measure rests on the assumption that the longer the delay between stimulus presentation and voice onset, the more processing is involved, and hence, the more effort was expended. Thus, over many reference turns, VOT should decrease asymptotically to certain (as-yet-undetermined) limits on linguistic processing.

This measure was computed automatically by the experimental software. The algorithm used a “moving window” technique to locate the first point in the file at which a certain percentage of the samples, or “hits”, fell above a threshold amplitude (computed as a percentage of the possible range of values). The size of the moving window was
fixed at a reasonably small value. Through pilot research, it was determined that a
window of 200 ms, with 2.27% of samples above a threshold of 11.44% of maximum
amplitude\textsuperscript{11} was optimal, given the recording level of the microphone.

After recording data for all participants, trials with abnormally short or long
VOTs were checked by hand. Though the algorithm was very accurate, some VOTs had
to be recomputed because of mouth clicks or accidental noises made by the director,
addressee (and alas, by the experimenter himself) prior to initiating the description.
These new VOTs were located by hand, using software designed for this purpose.

While the nominal measures provide a clear-cut test of the partner-specificity
of conventions, the scalar measures provide a more in-depth look at how directors use
prior and local context in formulating descriptions. The scalar measures, unlike the
nominal measures, can illuminate the relative contribution of prior and local context to
directors’ pre-verbal plans and post-VOT monitoring. This is because it is much easier
with continuous dimensions to systematically manipulate the stable and variable aspects
of context. Including VOT, there were a total of four dependent measures used in the
scalar trials.

**Conventionalization.** The scalar probe and transfer trials were transcribed and
coded for the presence of scalar adjectives classifying a target along two continuous
dimensions, size and shade (e.g., “small / little”, “large / big”, and “dark / light”). Recall
that throughout the training trials, each of the two kinds of scalar targets (circles and

\textsuperscript{11} For the audio data recorded in this experiment, this value was ±3750, approximately 11.44% of
the maximum amplitude value for 16-bit audio data (±32,768).
squares) was paired with one and only one kind of contrast object, such that, for example, it would be necessary to refer to circles by size and squares by shade. Over time, directors would conventionalize these scalars to refer to particular values along these continuous dimensions. Thus, a particular size of circle would be conventionalized as the “small” circle, and a particular shade of square as the “dark” square. While a phrase such as “small circle”, should initially reflect the relative size of a circle within a given stimulus array (local context), over time it would come to pick out a particular size of circle from the set of all circles that appeared in the experiment (prior context).

The scalar probe and transfer trials allow the assessment of the conventionality of scalar use. Following Lewis, I consider the use of a scalar to reflect a “convention” inasmuch as it would, given the local context, override other descriptions which would be more salient. Specifically, a scalar is conventionalized in the degree to which it is used in spite of the fact that strict adherence to the Maxim of Quantity and characteristics of the local context would recommend other alternatives. In the probe trials and in the unique shape trials of the transfer phase, it would be sufficient to describe the target shape as either “the circle” or “the square”, because it is the only shape of its category in the array. Thus, uses of “small” or “dark” would violate the Maxim of Quantity with respect to the local context, and would be permissible only to the degree to which these terms represent mutually known conventions.

---

12 The non-conventional contrast items in the transfer phase were also used to assess degree of conventionality. Unlike the unique shape items, in the non-conventional items it was sufficient to merely refer to a target shape by its non-conventionalized dimension (e.g., “dark circle” or “small square”). Uses of the conventionalized dimension, therefore, are redundant with respect to local context.
The work of Deutsch and Pechmann (1982), however, raises a small caveat to this operationalization of convention. In their experiments, they noted that people tended to use scalars redundantly and thereby routinely violate the Maxim of Quantity. They were able to explain some of these uses through prior context, but nonetheless, strict adherence was not observed. This raises the issue of whether redundant usage is a good measure of conventionality. Fortunately, the first few probe trials provide a baseline of the degree to which directors tolerate redundancy in their descriptions. If, over the course of the experiment, the use of the conventionalized terms systematically increases significantly above baseline, then this indicates that these scalars are not merely due to the natural redundancy that Deutsch and Pechmann reported.

Employing this definition of conventionalization makes it possible to test the degree to which convention use is governed by mutual knowledge and optimal design versus the mechanisms of dynamic coordination. Under the mutual knowledge hypothesis, with the new addressee of the transfer phase communicators should decrease their convention use to baseline levels (as given by the first probe trials of block 1-2), because conventions are a form of partner-specific “conceptual pacts”. If directors seek to minimize collaborative effort, they should formulate their descriptions taking into account only information that is co-present with their new partner—in this case, physically co-present information. Otherwise, addressees might construe directors’ descriptions as a violation of the Maxim of Quantity.

In contrast, the theory of dynamic coordination predicts two possible outcomes, depending on the assumptions that are made about how adjustment is
prompted. If adjustment is feedback-driven, there should be no difference across the transition between the acquisition and transfer phases. A spontaneous, director-initiated monitoring process, which operates when sufficient attentional resources are available, predicts that convention use should decrease, but not so far that the rate would return to baseline levels.

**Fixation measures.** The directors’ point of gaze was monitored to examine the degree to which directors used local context in planning and monitoring their descriptions. In the extreme, when directors used only prior context in preparing a description, only one fixation would be observed, located on the target shape. Conversely, when directors used only the local contextual information in the visual display, many eye movements would be required. It is assumed that because of the extra processing required by eye movements, directors who wish to minimize effort should, over the course of many reference turns, strive to minimize their use of local context while maximizing their use of prior context.

Software was developed to automate the analysis of eye fixations. The software calculated fixations within the sixteen squares of the grid. A fixation was defined as six consecutive digital frames (100 ms) within a given square of the grid. In computing fixations, frames that were lost due to blinks were discounted. A fixation was coded as complete once the eye position was located outside of the square for six consecutive frames. The program automatically displayed the number of pre- and post-VOT fixations.
In addition to the overall number of fixations, the software also counted the number of different shapes that were fixated on. This measure is useful because it gives an indication of the amount of distinct contextual information considered by the director. Consider the possibility that eight pre-VOT fixations are observed in a given trial. Though it would be possible that the director fixated on eight of the 16 shapes in the grid, it is also possible that the director fixated on the same two shapes four times each. In short, the number of different shapes fixated on provides a somewhat independent, converging measure for the use of contextual information.

When the eye fixation measures are combined with VOT, they yield a powerful index of the differential use of prior and local context in the planning and articulation phases of utterance production. Those fixations that occur prior to VOT (henceforth pre-VOT fixations) reflect the gathering of information in order to formulate a preverbal plan. By the time directors begin to articulate preverbal plans, they have made certain commitments to lexical selection and phrasal structure. Therefore, any fixations that occur after VOT (henceforth post-VOT fixations) reflect directors’ attempts to monitor their verbal plans for adequacy.

This measure is the most powerful one for differentiating between the various hypotheses under consideration. If directors consider mutual knowledge when designing and utterance, then the pattern of fixations relative to VOT should differ across the acquisition and transfer phases. The number of pre-VOT fixations should be slightly elevated in the transfer phase relative to the last block of acquisition, because directors
would need to make greater use of physically co-present information in formulating their
descriptions.

In the extreme, if directors optimally design their utterances for the new
addressee, both VOT and the number of pre-VOT fixations should return to near baseline
levels (i.e., to the values observed in Blocks 1-2), while the number of post-VOT
fixations should remain constant. At the other extreme, if directors are egocentric and
continue using prior context at the same rate, then VOT and the number of pre-VOT
fixations should remain constant. The number of post-VOT fixations, however should
increase drastically, because these descriptions would not be appropriate for addressees
and would require adjustment.

Additionally, the number of pre-VOT fixations can be correlated with the
degree of conventionalization. This provides a critical test of the hypothesis that
communicators’ use of conventions is related to the amount of local contextual
information that is considered in formulating a preverbal plan. If the hypothesized
relationship exists, then the number of pre-VOT fixations should be negatively correlated
with the rate of convention use.

2.3 RESULTS AND DISCUSSION

The items used in this experiment are visual arrays whose properties are
known and were subject to strict experimental control. For these reasons, results are
reported for the subject analysis only. A recent paper by Raaijmakers, Schrijnemakers,
and Gremmen (1999) advocates the exclusion of the item analysis in such circumstances.
Based on Clark (1973), they argue that it is only necessary to control item variance statistically when randomly sampled materials are used and the item variance is not controlled through experimental means.

2.3.1 Nominal Trials

2.3.1.1 Description Length and Lexical Diversity

Because each subject contributed only two data points per block to the nominal trials, the blocks were grouped in sets of two for purposes of analysis. Figure 1 charts the description length by block and condition.

   The baseline description length of 8.07 words in blocks 1-2 was observed to decrease by 3.59 to 4.48 words in blocks 7-8 ($t(15) = 4.28, p < 0.05$). The log number of words during the acquisition phase was submitted to a linear trend analysis, which revealed a significant trend, $F(1, 15) = 19.82, MSe = 0.176, p < 0.05$. However, as both theories predict, there was a significant increase in the length of descriptions in the beginning of the transfer phase. The description length increased from an average of 4.48 words in blocks 7 and 8 to 6.80 words in the transfer phase, $t(15) = 5.53, p < .05$. Nonetheless, as predicted by dynamic coordination, this value was 1.27 words less than the baseline length of 8.18 from blocks 1 – 2, a difference which was also significant ($t(15) = 1.78, p < .05$).

---

13 The significance levels of $t$ reported throughout are for a one-tailed distribution.
These results appear to favor dynamic coordination, inasmuch as the average length in the transfer phase is significantly less than the baseline value and greater than the length at the end of the acquisition phase. This suggests that communicators were at least partially recycling verbal plans. However, it is unclear whether these descriptions, which were only slightly more than one word shorter than baseline, would have been any less optimal for addressees.\textsuperscript{14} In other words, it is not clear whether there is “insufficient adjustment” as would be predicted by egocentric anchoring and adjustment. Nonetheless,

\textsuperscript{14} In only 2 trials (1.6\%) of the first four transfer items did addressees actually select the wrong referent.
it demonstrates that mutual knowledge is not wholly responsible for the decrease in description length over repeated reference turns.

Less critical to the hypotheses, but still of interest, is the lexical diversity of directors’ descriptions. This measure is depicted in Figure 2.9. This measure simply confirms that the description length was not simply due to the shortening of descriptions, but also to the process of codification, providing further evidence that directors were recycling their verbal plans. Like the number of words, there was also a significantly decreasing trend in the log-transformed values, $F(1, 15) = 20.09, MSe = 0.202, p < 0.05$.

As evident in the figure, directors' descriptions quickly approach asymptote, though some variability remains until the end of the acquisition phase (blocks 7 – 8). In blocks 3 – 4, only 20% of the words in the description had not been used in the previous description. By blocks 7 – 8 of the acquisition phase, 10% of the words were still new. It is important to keep in mind that the nominal trials were separated by between 17 – 20 trials. If it is assumed that directors strive toward uniformity in their descriptions, this variability could reflect degradation in the memory trace of the verbatim form due to the intervening trials.
Figure 2.9. Lexical Diversity of Descriptions of Nominal Shapes

As with description length, as directors switched partners for the transfer phase, their descriptions became more diverse. However, the lack of an adequate baseline precludes the possibility of determining whether this adjustment is insufficient. Though these measures indicate that directors do, in fact, make adjustments when they are faced with new addressees, it does not tell us why they would do so, nor when these adjustments occurred. First, it is possible that directors spontaneously take into account the identity of the addressee when formulating a preverbal plan. Second, although they may not consider the identity of the addressee when conceptualizing a message, they may spontaneously engage in monitoring and adjustment during
formulation and articulation of the utterance plan, as Horton and Keysar (1996) suggest. Both of these alternatives assume that adjustment is initiated spontaneously by the director.

An alternative is that these adjustments are actually prompted by feedback from the addressee. Through back channels (Yngve, 1970; Duncan, 1973), such as pauses, nods, or other meta-communicative devices, directors can gain important information regarding how their addressee understands a particular description. In this experiment, the director has “on-line” feedback of the addressees understanding of the descriptions by monitoring the position of the computer mouse. If, after the director issues a description, the mouse should stay put for an undue period of time or begin to move toward a contextual shape instead of the target, the director would realize that the description was either insufficient or misleading. By taking advantage of feedback, the director could opportunistically amend the description as necessary to avoid being misunderstood.

How is it possible to determine if an adjustment is spontaneously initiated by the director or prompted by addressee feedback? Unfortunately, the idea of tracking the position of the mouse did not occur to me until well into data collection. Thus, it is not possible to definitively decide between director-initiated monitoring and feedback-driven correction for the nominal trials. However, it is still possible to assess with less definitive means which aspects of a description are probably recycled components of previous utterance plans, and which are post hoc amendments.
It was possible to break each description along its “seams” into different clauses or tone units. The seams could be identified through pauses, descending intonation, or through recycled phrases. The early clauses are more likely to be part of a recycled verbal plan, while later clauses are more likely to be amendments. To get an estimate of the length of the preverbal plan, it is possible to consider only the description length of the base clause. For instance, one director in the transfer phase produced the following description of the shape on the left-hand side of Figure 2.2: “The peg-like thing .. looks like a parking place, a parking spot”. In this case, the base description is “the peg-like thing” which consists only of three words. ¹⁵ In contrast, the description, taken as a whole, is 11 words long. In another case, a director produced the description: “Uh, the square pie. Is the square with things cut out of it.. the uh weird shape, that's the square pie. Nope, nope, that one.” The break between the first phase (“pie” and “is”), while not at a traditional sentence boundary, was clearly marked by descending intonation on the word “pie” followed by a pause. The length of the base description for this instance is three words. Non-content “filler” words such as “like”, “sort of”, “kinda” were also omitted from this analysis.

In this way, the base description lengths for the transfer phase were computed. These descriptions were of an average length of 5.58 words, significantly longer than the base length of 4.14 in blocks 7-8 ($t(15) = 4.08, p < 0.05$). Thus, of the 2.32 word increase in overall length from the acquisition to the transfer phase, the analysis indicates that

¹⁵ Some might argue that “the peg-like thing” is actually four words. This is a matter of personal taste and belief in prescriptive grammar. I consistently coded hyphenated words as one word.
roughly half of these words might have been from recycled plans, while the other half were probably prompted by addressee feedback.

2.3.1.2 Voice Onset Time

Assuming that longer descriptions take longer to plan, a better way of testing hypotheses about when amendments occur is to examine voice onset time. The results of the analysis are presented in Figure 2.10.

![Figure 2.10. Voice Onset Time by Block and Condition.](image)

It is assumed that the longer and more elaborate the description, the longer it should take to prepare. Observing that description length increased from the end of the acquisition phase to the beginning of the transfer phase, if these increases were part of the verbal plan, then there should have been a concomitant increase in VOTs. As the figure
makes clear, VOT actually continued to decrease from the acquisition to the transfer phase. This provides strong evidence that whatever additional content was added to directors’ descriptions was not part of the verbal plan, but was added post hoc.

One problem with this analysis is that it doesn’t take into account the possibility that directors are actually beginning their descriptions in the transfer phase with hedges such as “um” and “uh”, because they realize that with the new partner they will need to engage in extra planning. In fact, 33% of the nominal trials in Transfer 1-2 phase began with a hedge such as “um” or “uh”, or included an early pause, often after uses of the definite and indefinite article, as in “it’s a...”, or “the... uh...”. This is compared to only 27% of the trials in Acquisition 7-8. Because the VOT algorithm does not discount these pauses, this could mask any extra delay in voice onset.

To control for the possibility that VOT times were deflated by hedges, I returned to the original sound files for the end of the acquisition phase (blocks 7 – 8) and the first four nominal trials of the transfer phase, and calculated VOT by hand, excluding phrase initial pauses. Thus, for a description such as “uh... it’s the... small peg thing”, VOT would be coded as beginning at the “s” in small.

The new VOT values were 1204 ms and 1150 for the end of the acquisition and beginning of transfer, respectively. The difference was still in the opposite direction of what would be predicted by the Initial Design model.

Taken together, these findings suggest that although directors do adjust their descriptions when confronted with a new partner, it is likely that this adjustment doesn’t figure into the preverbal plan but is driven by either spontaneous, self-initiated
monitoring posterior to voice onset, or other-initiated feedback signals. This is suggested principally by the fact that VOT does not substantially increase when describing an old referent to a new addressee. In other words, directors appeared to be recycling verbal plans, and refashioning them when prompted by multimodal feedback.

The fact that directors attempt to adjust these descriptions on-line instead of during utterance planning can account for the “installment-like” nature of some directors’ descriptions in the transfer phase, such as “the backwards C... it has something sticking out of it on the right side” or “uh .. the round pie. It's the circle with pieces cut out.” It is likely that the process of conventionalization creates linguistic products, such as “the backward C” or “the round pie” which are difficult for directors to ignore when producing descriptions for new addressees because they are easily retrieved immediately from memory.

2.3.2 Scalar Trials

Though the nominal trials provide a clear, general test of different models of partner adaptation, they cannot illuminate the use of context in how these descriptions were formulated, because nominal conventions are defined across discrete category boundaries. For nominal trials, it is not readily apparent how to operationalize the use of prior and local context. For instance, when a director opts to call a particular shape “the tetris shape”, how much of that choice is based upon the similarity / dissimilarity of that shape with other shapes in the local context, and how much with prior context (whatever that may be)? The answer is not obvious. The scalar trials, in contrast, can illuminate the
mechanisms behind partner adaptation and effort minimization because the dimensions along which shapes contrast with one another are continuous, well-defined and easily manipulable.

2.3.2.1 Rate of Convention Use

The first measure reported here amounts to a simple validation of the assumptions behind the scalar trials. This measure is the rate that scalar terms were used conventionally over various reference turns. Directors are said to use a scalar term conventionally when they reliably use it to describe a target shape, even when that shape is not accompanied by a contrast object. Thus, “large circle” would conventionally come to refer to a circle of a particular size, and would be used even when there is no smaller circle in the visual array. The conventional use of a scalar, then, relies exclusively on prior instead of local context.
Figure 2.11. Rate of Scalar Convention Use by Phase and Block.

Figure 2.11 shows, by phase and block, the percent of probe trials in which the conventionalized scalar was used. During the acquisition phase, convention use rose 14 points from its baseline level of 33% to 47%. A linear trend analysis performed on the arcsine-transformed proportions revealed a significantly increasing trend, \((F(1, 15) = 13.39, MSe = 0.013, p < 0.05)\).

Turning to the critical hypothesis of this study: did directors reduce their use of conventions in the transfer trial when facing new addressees? Initially, they did not. From acquisition blocks 7-8 to the first four scalar trials of the transfer phase, directors’ convention use only dropped 3%, from 47% to 44%. These values, subject to an arcsine-transformation, were not significantly different from each other \((t(15) = 0.35, \text{n.s.})\). However, convention use in the first four trials of transfer was still marginally different
from the baseline level \((t(15) = 1.44, p < 0.09)\). Interestingly, over the course of the
transfer trials, the conventionalization rate dropped from 44% back to the baseline level
of 28% \((t(15) = 1.71, p < 0.06)\), supporting the idea that directors dynamically adjusted
their use of context dynamically in response to negative feedback from the addressee (see
Fig. 2.12).

According to the mutual knowledge hypothesis, a director should rely on prior
context in describing a shape only when the addressee shares this contextual knowledge.
From the new addressee’s perspective, these conventions would appear as redundant and
“overinformative”. They violate both the Principle of Optimal Design and the Maxim of
Quantity, because they presuppose that a relevant contrast set exists.

Figure 2.12. Rate of Convention Use Over Transfer Trials.
Figure 2.13. Sample Transfer Trial from Experiment 1.

A vivid example of how this redundancy impairs comprehension comes from a transfer trial in this experiment, in which a director described a target square to a new addressee. The experimental item is represented in Figure 2.13. In the figure, the target shape appears in the bottom left corner. The director had previously conventionalized this particular shade of square as “the light square” during the acquisition phase. When describing the shape the first time for the new addressee, the director referred to it as “the light square” even though there was no dark square with the same orientation in the array to provide a contrast. The addressee selected the lighter of the two squares in non-canonical orientation as the target (most of the directors referred to these shapes as “diamonds”).
The addressee’s behavior makes sense if we consider the fact that the use of the scalar “light” in describing a square presupposes the existence of a contrast in shade between at least two squares. However, as this addressee conceptualized the display, there were actually three squares; the square in the canonical orientation (which was the target square for the director), and a light and a dark square rotated at a 45 degree angle. Because the only contrast between two shapes that existed in the display was between the two squares in non-canonical orientation, the addressee chose the lighter of the two. Thus, the pair made an error because the speaker violated the Maxim of Quantity. This demonstrates that the egocentric behavior of speakers can lead to systematic errors.

2.3.2.2 Voice Onset Time

![Voice Onset Time Graph](image)

Figure 2.14. Voice Onset Time.
VOT dropped from its baseline level of 1232 ms to 1062 ms in Blocks 7-8. The linear trend was significant \((F(1, 15) = 5.92, MSe = 39247, p < 0.05)\). Assuming that VOT is a measure of the amount of planning that goes into an utterance, the linear decrease in VOT over the acquisition phase suggests that directors were, in fact, increasing their use of prior and decreasing their use of local context in producing their descriptions. The mutual knowledge hypothesis predicts that there should be an increase in VOT in the transfer phase, because they would have to revert to the local (physically copresent) context when preparing their descriptions. However, VOT actually continued to decrease from 1062 ms to 950 ms. The fact that it did not significantly increase is consistent with the observation that directors continued to use the scalar conventions at the same rate in the beginning of the transfer phase as they did at the end of the acquisition phase.

Once again, the objection could be raised that perhaps directors in the acquisition phase had to engage in more planning, and were therefore beginning their descriptions with pause signals such as “um” or “uh”. However, only 10% of descriptions in the acquisition phase began with pause markers, which was no different from the proportion of pause markers in acquisition blocks 7-8. This makes such an interpretation implausible.

It is possible that the decrease in VOT simply reflects greater efficiency in utterance planning, and not decreased use of local context. The number of pre- and post-VOT fixations provides an important check on this measure.
2.3.2.3 Fixation measures

This section reports the results for two different converging measures that were computed relative to VOT: number of overall fixations and number of distinct shapes that were fixated upon before and after VOT.\(^\text{16}\)

The two different measures are related to different aspects of utterance production. It is assumed that the number of pre-VOT fixations is related to the formulation of a pre-verbal plan, while the number of post-VOT fixations is related to monitoring processes. Both measures provide powerful insight into the use of contextual information in language processing, and allow the testing of the two theories of coordination.

If conventions minimize speaker effort, then the number of pre-VOT fixations (both overall and on distinct shapes) should drop over the course of the acquisition phase. Once directors face new addressees, however, dynamic coordination predicts that there should be no difference in the number of fixations for the first few items. The Initial Design model predicts that there should be a rise in the number of pre-VOT fixations, as directors attempt to tailor their descriptions with respect to their mutual knowledge with the addressee. This is because they would have to heed the local, physically copresent context and could no longer count on the accumulated prior context.

\(^{16}\) Unfortunately, the eye data for two participants turned out to be unusable because of mechanical problems with the eyetracking equipment. Software crashes during the experiment made the data for these two participants unrecoverable.
These two models also make different predictions about changes in the number of post-VOT fixations. If directors formulate their descriptions following the principle of optimal design, then the number of post-VOT fixations should not change from the acquisition to the transfer phase. However, if directors’ preverbal plans are formulated egocentrically, then they might seek out information that their plans were adequately understood. Furthermore, addressees would have difficulties understanding these descriptions, and the number of post-VOT fixations should show a dramatic increase.

Figure 2.15 displays the results for the overall number of pre- and post-VOT fixations. The baseline number of pre-VOT fixations in Acquisition 1-2 was 4.25, which
dropped by 0.79 points to 3.46 in Blocks 7-8. This decrease was marginally significant $(t(13) = 1.46, p < 0.09)$, and the overall linear trend was in the right direction, but not significant $(F(1, 13) = 2.63, p > 0.12)$. When directors confronted new addressees in the transfer phase, the number of pre-VOT fixations remained virtually identical (3.47) to the end of the acquisition phase (3.46). This result is predicted by dynamic coordination, and provides strong evidence that directors did not take into account the change in mutual knowledge when they designed their utterances for new addressees.

Even more compelling evidence for dynamic coordination is the fact that while the number of pre-VOT fixations remained relatively constant after the addressee switch, the number of post-VOT fixations skyrocketed. In Acquisition 7-8, the average number of post-VOT fixations was 4.71. After switching partners, it shot up to 6.49 fixations $(t(13) = 4.23, p < 0.05)$. Thus, whatever partner adjustment occurred from the Acquisition to the Transfer phase, it would appear that 100% of it took place after directors had already formulated and began articulating preverbal plans. Instead of scanning the context to take into account the mutually known physically-copresent information and integrating this information into their preverbal plans, they instead violated the normative standards of the Maxim of Quantity and the Principle of Optimal Design and behaved as if the conventions were transparent. The dramatic rise in post-VOT fixations suggests that they also engaged in more monitoring with the new addressee than they did with the old addressee in the last blocks of the acquisition phase. However, the question arises as to whether directors spontaneously increased their
vigilance of how the addressee understood the message, or whether the monitoring was simply prompted by signals of confusion from the addressee.

The idea that these post-VOT fixations were driven more by addressee feedback than spontaneous monitoring processes is supported by the observed decrease in the number of post-VOT fixations during the acquisition phase. The number of fixations decreased from a baseline value of 6.20 to an average of 4.71 fixations immediately prior to the transfer phase. The linear trend was reliable ($F(1, 13) = 7.09, p < 0.05$). If directors were spontaneously initiating monitoring processes, then there should have been a rise instead of a decrease in the number of post-VOT fixations. To see why, consider the fact that speakers produced utterances with greater efficiency over the course of the acquisition phase, as shown in the decline in the length of VOT and the number of fixations prior to VOT. The greater efficiency of processing should have freed up extra cognitive resources that could have been channeled toward utterance monitoring. The fact that they did not do this spontaneously suggests that they only monitored their utterances when they sensed that the addressee was having trouble locating a referent.

Further support for these conclusions is provided by the other fixation measure, the number of distinct shapes on which the director fixated. This measure is shown in Figure 2.16. The number of distinct shapes fixated prior to VOT dropped from 3.71 to 3.05 during acquisition. This decreasing linear trend was significant, $F(1, 13) = 4.96, MSe = 0.738, p < 0.05$. In the transition from the acquisition to the transfer phase, the number of pre-VOT shapes went from 3.05 to 3.13. This slight increase of 0.06 shapes was not significant ($t(13) = 0.50, p > 0.05$). A significantly decreasing linear
trend in the number of post-VOT shapes was also observed ($F(1, 13) = 7.37, MSe = 0.560, p < 0.05$), decreasing from 4.54 to 3.80 in the acquisition phase. Unlike the number of pre-VOT shapes, from acquisition to transfer the number of post-VOT shapes showed a 27% increase, rising to 4.84, $t(13) = 3.74, p < 0.05$. As with the overall number of fixations, the number of shapes supported the dynamic coordination model. By comparing Figures 2.15 and 2.16, the same pattern can be seen in both fixation measures.

**Figure 2.16. Number of Distinct Shapes Fixed, Pre- and Post-VOT**

In the interpretation of these measures, the assumption of a relationship between pre-VOT fixations and preverbal plans has been critical to testing the claims of the model. It is important, then, to consider a measure which directly establishes the link
between these two factors. If the assumption is true that the number of fixations is related to the use of local versus prior context, then the number of fixated pre-VOT shapes should be negatively correlated with the use of the convention. This means that the less that directors consider local context, the more likely they should be to use the convention.

![Figure 2.17. Number of Pre-VOT Fixations Versus Rate of Convention Use](image)

Figure 2.17. Number of Pre-VOT Fixations Versus Rate of Convention Use

The convention rate for each participant was correlated with that subject’s average number of pre-VOT fixations. The analysis produced a correlation coefficient of \( r = -0.61 \) (\( p < 0.05 \)). However, upon viewing the scatterplot for each participant it was appeared that two participants may have disproportionately affected the correlation in
favor of a strong negative relationship. Rather than delete the participants who showed
the correlation most strongly, it was determined to divide each subject’s data in half,
thereby somewhat diluting the effect of these participants on the overall analysis. Even
then, the coefficient was -0.57 (p < 0.05). A scatterplot with a regression line is shown in
Figure 2.17. The number of pre-VOT shapes was found to explain 32% of the variation
in rate of convention use. This unambiguously demonstrates a relation between these two
measures. The more exhaustively that directors looked around when establishing a
preverbal plan, the less likely these plans were to include the convention.

One might wonder whether such a direct link can be established between post-
VOT fixations and monitoring processes. As noted above, the decline in post-VOT
fixations during the acquisition phase supports the idea that much of dynamic adjustment
is prompted by addressee feedback.

During the transfer phase, directors appeared to alter how they used context in
planning and adjusting descriptions of target objects. The number of pre- and post- VOT
fixations is shown in Figure 2.18. The number of pre-VOT fixations gradually rose over
the course of transfer, rather than all at once as suggested by mutual knowledge
(marginally significant linear trend, $F(1, 13) = 4.21, MSe = 1.63, p < 0.07$). This is
consistent with the decrease in the rate of convention use represented in Figure 2.12.

To the degree to which monitoring is director-initiated, the number of post-
VOT fixations on contextual shapes should have remained fairly constant over the course
of the acquisition and transfer phases. In fact, it could be argued that the number of post-
VOT fixations should have risen during acquisition, as greater processing efficiency freed
up resources for monitoring. To the degree to which it is initiated by addressee feedback, post-VOT fixations should decrease over the acquisition phase (as the addressee adapts to the director’s descriptions) and increase sharply in the transfer phase, when the director confronts a new addressee. The increase in the transfer phase would be prompted by the difficulties new addressees would have in interpreting the directors’ descriptions. For instance, if the director referred to the target square in Figure 2.13 as “the light square”, and then observed the mouse pointer moving toward the lighter of the two diamonds, the director could then pre-empt this mistake by providing a correction such as, “the only square”. As the speaker follows the mouse pointer, this would incur more fixations.

Figure 2.18. Number of Pre- and Post- VOT Fixations in Transfer Phase
In the first four trials of the transfer phase, the number of post-VOT fixations showed a parabolic trend in the transfer phase $F(1, 13) = 7.27, p < 0.05$, reaching a peak in Transfer 3 of over 9 fixations! The average number of fixations in Transfer 3-4 was well-above the baseline number of fixations of 6.20 in Acquisition 1-2 ($t(15) = 2.90, p < 0.05$). This suggests that the conventionalization process caused difficulties in adjusting to the perspective of the new addressee. The number of fixations then dropped by the end of the phase to an average of 6.86. Because the number of post-VOT fixations did not remain constant, it would appear that this measure is more related to addressee feedback than to spontaneous, director-initiated monitoring.

### 2.4 CONCLUSIONS

The experiment sought to show that conventions are established and used independently of mutual knowledge, through the mechanisms of dynamic coordination. The results strongly support this view.

For the nominal trials, description length decreased according to an exponential decay function, similar to that found by Clark and Wilkes-Gibbs (1986). This is predicted by both theories. Yet once communicators faced a new addressee, though description length increased, it did not return to its baseline value at the beginning of the acquisition phase. It is important to note that both theories predicted an increase, though for different reasons. Mutual knowledge predicted an increase because communicators lacked task-relevant mutual knowledge with the new addressee.
Dynamic coordination predicted an increase due to the post-VOT refashioning of an inadequate description.

An analysis of the base description lengths suggested that communicators were at least partially recycling conventionalized aspects of their verbal plans. The fact that VOT failed to increase in the beginning of the transfer phase also suggests that communicators were recycling old plans and dynamically apportioning effort after VOT to secure reference. A further analysis also ruled out an alternative explanation of the results, according to which the null effect in VOT was an artifact of the increased number of “um”s and “uh”s in the beginning of the transfer phase.

Though the nominal trials support aspects of dynamic coordination, they are limited in the sense that they cannot provide evidence for the idea that communicators were making more use of prior and less use of local context over repeated reference turns. This is because it is hard to distinguish between the use of prior and local contexts with nominal shapes. In this experiment, the scalar trials performed this important function.

Dynamic coordination posits that communicators seek to minimize effort by maximizing their use of prior context, inasmuch as it does not compromise communicative success. An analysis of the percent of probe trials in which the conventionalized scalar was used revealed that communicators increasingly made use of prior context in their descriptions. Yet once they faced the new addressee in the transfer phase, mutual knowledge theory predicted that they would stop using these conventionalized scalars because they were not mutually known. Instead, it predicts that communicators should base their descriptions primarily on physically copresent
information. If they use the conventions instead, their descriptions would appear redundant and might be difficult for addressees to interpret. However, results indicated that they used the convention at the same rate as they did with the previous addressee.

While the use of prior context increased in the acquisition phase, the use of local context, as evidenced by VOT and the fixation measures, significantly decreased. By the end of the acquisition phase, it could have been possible for communicators to “max out” their use of prior context by simply using the convention (e.g., “small circle”) and completely ignoring the shapes in the local context. Such a description would have succeeded on 100% of the acquisition trials. However, communicators did not do this—though the average number of shapes they looked at before VOT steadily declined, they still looked at an average of about 2.5 shapes other than the baseline in Acquisition 7-8. This indicates that communicators maintain a minimal amount of vigilance of local context so that, should it suddenly change, as in the extreme shape transfer items, they could recover in time to prevent an error. Thus, there appears to be some constraint on the degree to which they attempt to maximize their use of prior context.

Furthermore, according to the normative claims of mutual knowledge, when the communicators faced new addressees in the transfer phase, they should have made greater use of the mutually known local information, as would be demonstrated by a rise in VOT and in the number of pre-VOT fixations. However, the number of pre-VOT fixations was identical from Acquisition 7-8 to Transfer 1-2, and the difference in VOT was in the wrong direction. These results strongly refute the predictions of the Initial Design version of mutual knowledge theory.
It is interesting to note that by the end of the transfer phase, communicators use of prior context gradually returned to baseline. In other words, it gradually came to match the mutual knowledge that they shared with the addressee, but this was clearly a result of dynamically re-weighting the meanings of the scalar terms based on negative feedback from the addressee, and not initial design.

A final important validation was provided for the theory of dynamic coordination through the examination of the relationship between the use of local context and rate of convention use. The dynamic coordination model predicted that the number of shapes that the communicator fixated upon prior to VOT would be negatively related to the rate of conventionalization. In support of dynamic coordination, the fixation measure predicted 33% of the variance in the use of the conventionalized scalar. This provides further confirmatory evidence for the dynamic coordination theory.

These results provide the strongest evidence yet against the “Initial Design” model of utterance production. Whereas the communicators in Horton and Keysar’s experiments showed some use of mutual knowledge when they were not placed under a deadline, the communicators in this experiment showed no effect whatsoever, even though natural conditions of production prevailed. Polichak and Gerrig (1999) challenged Horton and Keysar’s conclusions that mutual knowledge is used in adjustment and not in initial design, because Horton and Keysar imposed an artificial deadline. However, no such argument could apply here—the communicators interacted with real addressees, and were allowed to formulate descriptions at their leisure.
One fascinating observation that arose during debriefing was that many communicators reported not explicitly noticing the exact nature of the size/shade pairings in the training items, e.g. that target circles always appeared with circles of a different size. It is possible, then, that the products of conventionalization through dynamic weighting and the use of prior context are implicit and therefore not accessible to conscious control. Because of this, the communicators in this experiment behaved as though they experienced an “illusory transparency of convention”—they appeared to believe these conventions would be understood even when they could not infer mutual knowledge of them through community membership. In closing, the results from this experiment refute Clark and Carlson’s (1981) claim that mutual knowledge is an essential condition for convention use. In addition to mutual knowledge, conventions can come about through dynamic coordination.
CHAPTER THREE

CONVENTION USE IN LANGUAGE COMPREHENSION

3.1 INTRODUCTION

In Experiment 1, when communicators faced new addressees in the transfer phase, they formulated utterances egocentrically, using conventions even when they were not mutually known, and then adjusted their descriptions based on communicative success and addressee feedback. In doing this, they appear to have strayed far from the normative standards of the Maxim of Quantity and mutual knowledge, principles that suggest that they should design their utterances “optimally”; that is, with respect to the physically co-present information that was mutual known with the new addressee.

The surprising egocentrism that communicators exhibited suggests that they strive to minimize not collaborative, but their own effort. This raises many questions regarding how addressees comprehend communicators’ egocentric referential expressions. Do they assume that descriptions have been optimally designed for them? If not, then how do they locate referents? Do they use mutual knowledge, or are they egocentric? Do addressees establish “conventions” for interpreting a particular communicators’ descriptions? The experiment described in this chapter seeks to answer these questions.
3.1.1 Mutual Knowledge as the Context for Comprehension

What kinds of contextual information do addressees use in understanding communicators’ utterances? Clark and Carlson (1981) argue that the traditional notion of “context” has two important drawbacks. First, it is simply too vague to be useful. It can refer to information which is immediately available in perception, accessible in working memory, or even the whole of a person’s background knowledge. Second, what is traditionally accepted as “context” is untenably broad, making it psychologically implausible that addressees could retrieve and use contextual information efficiently enough for conversation. To overcome these drawbacks, they propose that the context for comprehension consists of the set of mutually held assumptions, beliefs, and knowledge that constitute an addressee’s “common ground” with a speaker. They claim that the comprehension system “must distinguish between information that is and is not part of common ground, because otherwise in certain situations it will systematically misinterpret conventions, direct and indirect speech acts, definite reference, and contextual expressions.” (p. 76). Clark and Carlson argue that when addressees restrict the information they consult their common ground with speakers, they streamline memory search, making comprehension less effortful. The comprehension system can be said to be “optimally designed” to the degree to which it limits information search to

---


18 Not to be confused, however, with the Principle of Optimal Design.
common ground. Clark and Carlson concede that, “Whether its design is actually optimal in this respect is a question that can only be answered empirically” (p. 76).

The idea that addressees restrict the knowledge they consider to mutual knowledge can be termed the hypothesis of restricted search. Clark, Schreuder, and Buttrick (1983) sought evidence for this hypothesis. They argued that in understanding instances of demonstrative reference—utterances such as “this flower”—addressees select referents that are mutually salient with respect to common ground. In their study, they approached members of the Stanford University community and showed them one of two pictures of a flower display. One picture showed a floral display with flowers that were all roughly of equal salience, and the other, a single yellow flower that stood out from all of the other flowers in the display. When they asked addressees, “How would you describe the color of this flower?”, addressees were more likely to ask “which one?” or “this one?” when they viewed the first picture than when they viewed the second. In the latter case, the most common response was simply, “yellow”. Clark, Schreuder, and Buttrick interpreted this as evidence that addressees were using the physical co-presence heuristic to select the most salient flower with respect to common ground.

Keysar (1997) analyzed the design of this study, along with several other key papers that sought to test hypotheses related to restricted search. He found that the design of these studies systematically confounded common ground with perceptual or attentional salience. For instance, participants in the flower study could have simply based their responses on perceptually salient information instead of common ground. When there was a perceptually salient referent, they would simply describe its color. When none of
the referents was more salient than the others, they would ask for clarification. Keysar pointed out that although the studies reported by Clark, Schreuder, and Buttrick demonstrated that addressees used an available piece of information to comprehend utterances, they did not show that addressees used this information because it was mutually shared.

In fact, Keysar (1994) has shown that even when people have informed and accurate reference diaries—when they are certain about who knows what—they don’t always use them in communication. He discovered that when readers assess how an addressee would understand an utterance, they often rely on information that they know is inaccessible to the addressee. When readers know the intention of a speaker, they tend to believe that addressees would perceive that intention, even when they know that the addressee is missing crucial information. Keysar termed this phenomenon “the illusory transparency of intention”, because it suggests that people believe that it is possible to “see through” the linguistic form of an utterance to its intended meaning.

3.1.2 Perspective Adjustment

Though many researchers have sought evidence for or against the idea that addressees use common ground in comprehension, until recently, no one had examined exactly how they use that information. Recently, Keysar and I conducted a series of eyetracking studies of spoken language comprehension that sought to fill this void (Keysar, Barr, Balin, & Brauner, in press; Keysar, Barr, Balin, & Paek, 1998). The results of these studies challenge two claims: (1) that common ground is the context for
comprehension, and (2) that a comprehension system thus designed would, in fact, be “optimal”. We found that addressees do not appear to initially interpret speakers’ utterances with respect to common ground, but instead, use information that is salient to them, even when they know that this information is inaccessible to the speaker. If they use common ground at all, they appear to use it mainly to monitor and adjust the initial products of comprehension.

Keysar, Barr, Balin, and Brauner (in press) designed a referential communication task that investigated how addressees comprehend speakers’ referential descriptions in face-to-face environments. In this experiment, addressees interpreted a confederate speaker’s instructions to move objects around in a four by four grid (see Figure 3.1). It was possible for both the addressee and speaker to view some of the objects in the grid, while other objects were occluded from the speaker’s view. In each grid, one of the instructions was ambiguous, such that if the addressees considered mutual knowledge, they would select a different referent than if they interpreted the descriptions with respect to perceptual salience alone.

In order to investigate the kind of information addressees considered in comprehending referential descriptions, we their eye movements as they performed the task. Previous work demonstrates that addressees fixate their gaze on objects when they consider them as possible referents (Eberhard, Spivey-Knowlton, Sedivy, & Tanenhaus, 1995), before they reach for these objects. This allowed us to examine whether addressees restricted their search for referents to mutually visible objects.
Figure 3.1 shows an example of a grid we used in this experiment. For this item, the speaker instructed the addressee to “move the small candle below the truck”. In this figure, a small, medium, and large candle are visible to the addressee. However, only the two larger candles are visible to the speaker. To the degree to which addressees interpret “the small candle” with respect to common ground, their eyes should fixate upon the smaller of the two larger, physically co-present candles. In contrast, to the degree to which they use perceptually salient information, they should consider the smallest, hidden candle as a referent.

![Addressee's View](image1.png) ![Director's View](image2.png)

**Figure 3.1. Sample item from Keysar, Barr, Balin and Brauner (in press).**

We compared the time course of fixations on the smallest candle in this test condition to a control condition in which the smallest candle was replaced by a control
object that was not a good referent for the phrase “the small candle”. Figure 3.2 shows a timeline of fixations in both the control (top) and test (bottom) conditions. The results were clear. In the test condition, addressees tended to fixate on the occluded candle (the “distractor”, marked by a triangle in the figure) before they initially fixated on the target candle (gray circle). They tended to eventually select the target candle as the referent, as indicated by their final fixation (black circle). However, their selection was delayed by over 1 second in the test condition relative to the control. Surprisingly, in a quarter of the test trials, addressees conspicuously began to reach for the distractor, and actually proceeded to move it in 17% of all test trials. The observation that addressees considered hidden referents—even to the point of selecting them as the “intended” referent—negatively answers Clark and Carlson’s query about the optimality of the comprehension system. The comprehension system does not appear to be designed to limit information to common ground.

19 For this specific item, the control object was a small toy monkey.
We explained these results in terms of a comprehension version of the “perspective adjustment model”, based on the mechanism of egocentric anchoring and adjustment. According to the model, addressees initially interpret descriptions using information that is available to them, and use mutual knowledge to correct the products of comprehension for errors. This would lead them to systematically misinterpret speakers’ referential descriptions, especially when privileged referents\(^{20}\) are salient to them.

Why don’t addressees use their knowledge of what is commonly known, especially when egocentrism leads to systematic misunderstanding? According to Clark

---

\(^{20}\) Privileged referents are non-shared objects, such as the occluded candle in Figure 3.1.
and Carlson’s view, Keysar et al.’s findings provide evidence that the comprehension system is designed suboptimally, because it fails to use important contextual information that would effectively limit the size of context. Even more suboptimal is the finding that addressees systematically misinterpret speakers’ referential descriptions.

In contrast, perspective adjustment views the egocentric behavior of language users as optimal, not because it ensures fail-safe communication, but because it reduces uncertainty in a way that reflects an ideal compromise between cognitive efficiency and communicative success. The heuristic of egocentrism simplifies the task of producing and comprehending utterances by reducing the uncertainty in estimating another person’s perspective. Instead of mutual knowledge, egocentric interpretation may be the mechanism which, in Clark and Carlson’s terms, helps to “cut these gargantuan contexts down to size” (pg. 71). Because common ground is a kind of meta-knowledge that is cognitively effortful and time-consuming to access, people appear to be opportunistic about using it. This means that people use mutual knowledge only to monitor and adjust the products of utterance formulation and comprehension, but not when they formulate or comprehend utterances. This behavior is optimal because it minimizes the amount of time and mental energy spent in processing language, while still permitting successful communication—especially since people can typically monitor the success of communication through feedback from other interlocutors.

The perspective adjustment model offers a strong challenge to the mutual knowledge hypothesis for comprehension. However, perspective adjustment been mainly applied to cases of “one-shot” reference. In these cases, no precedent exists which could
guide addressees in their interpretation of speakers’ descriptions. When interlocutors encounter referents over repeated instances, it is possible that they simplify the processes of referring through the use of conventions. Dynamic coordination theory extends the perspective adjustment model to account for the use of conventions in comprehension.

3.1.3 Dynamic Coordination and the Use of Conventions in Comprehension

Though the mutual knowledge hypothesis offers no model that describes how addressees might establish and use conventions in comprehension, it is possible to derive one possible model from the principle of optimal design and the idea that interlocutors form conceptual pacts. According to the mutual knowledge hypothesis, conventions depend on mutual knowledge. Interlocutors use conventions because they are part of mutual knowledge. For addressees, interpreting a communicators’ description with respect to a convention would simplify referential search, because the conceptual pact between the interlocutors would reduce the set of possible candidates to mutually known referents.

By the principle of optimal design, addressees should assume that speakers tailor their utterances specifically to the mutual knowledge they share with the addressee. An important prediction can be derived from this principle. If addressees assume that speakers adhere to optimal design, then when they encounter a new speaker with whom they share no conceptual pacts, then they should uniformly abandon the old pacts they established with the previous speaker. This is because they should engage in “restricted search” and search among referents that are mutually known with the interlocutor.
However, as Clark and Brennan (1991) acknowledge, it is possible that interlocutors abandon these pacts only gradually. Yet if they are only abandoned in the case of negative feedback—i.e., because communicative success is compromised—then they would be no different from egocentric comprehension strategies that dynamically adapt to minimize effort and maximize communicative success. If this is the case, then the theory of dynamic coordination should be preferred as the more parsimonious alternative, because it reduces the phenomenon to known cognitive principles instead of postulating specialized, partner-specific “conceptual pacts”.

According to dynamic coordination, addressees could simplify referential search by using prior context to locate referents. Thus, when a certain size of circle is conventionally referred to as “small”, an addressee does not need to exhaustively check context to discover which, of all circles, is the smallest. Rather, upon discovering a circle of a particular size, the addressee can terminate referential search immediately. In this way, descriptions that make use of prior context limit the amount of new, local contextual information that needs to be considered in a given instance of reference.

Through the process of dynamic adaptation, addressees who interpret repeated descriptions from a particular speaker arrive at comprehension strategies that are elegantly calibrated to the communicators’ use of words. However, unlike conceptual pacts, these comprehension strategies are not partner-specific. The theory predicts that when these same addressees are paired with new partners, they should adjust their comprehension strategies only when these strategies fail.
3.2 METHODS

3.2.1 Task

The task used in this experiment was, in many respects, identical to the task of Experiment 1. The experiment employed the same referential communication setup. Addressees listened to directors’ descriptions of target shapes and attempted to identify the target based on the description. As in the previous experiment, there was an acquisition phase followed by a transfer phase, in which the participant encountered a new partner.

There were several important differences from the first experiment. The principal difference was that addressee, instead of the speaker, was the subject of data collection. Additionally, addressees listened to pre-recorded descriptions from absent instead of live speakers.

Pre-recorded materials were used in this experiment instead of live confederates because it greatly simplified the design and procedure, and eliminated some difficulties and possible confounds that live speakers could cause. Because fixation latencies were the primary dependent measure in this experiment, it was important to strictly control the interstimulus interval (ISI) between the presentation of the visual array and the initiation of the description. Live speech is problematic because it is very difficult to control. The use of two live speakers is even more problematic and likely to produce confounds. This is because it would be impossible to ensure that the timing and content of descriptions from both speakers would be commensurable. If the timings were
to differ systematically between the acquisition and transfer phases—as they would almost inevitably do, with two live speakers—then any differences in the fixation latencies could be attributed to these timings alone instead of the manipulation itself.

Because pre-recorded materials were used, three important steps were taken to enhance the realism of the experimental situation: First, participants were told that the recordings were from previous participants in the experiment. Second, the stimuli included filler shapes that speakers referred to over the course of various reference turns, shortening their descriptions each time (as the real, live speakers did in the first experiment). This reinforced the idea that the temporal sequence in which the items were recorded was preserved in playback. Finally, a sound file consisting of brief, incidental conversation between the experimenter and the speaker from the transfer phase was played just prior to the first item of the transfer phase. This was to convince addressees that they would hear this recording from the beginning, and that the transfer phase speaker would have no knowledge of the conventions. These modifications are described in more detail below in the sections on materials and procedure. No participants spontaneously mentioned any problems or noted anything unusual during debriefing, suggesting that the staging was effective.

Unlike Experiment 1, addressees were not given feedback about their performance during the task because it was incompatible with non-interactive nature of the communicative situation. Feedback in the non-interactive situation might induce addressees to adopt different strategies or adjust more than they would in a truly interactive situation, because in an interactive situation, both participants would share the
burden of any errors. In any case, the absence of feedback was inconsequential because participants made virtually no errors on the task.

Yet another major difference from Experiment 1 was in the number of shapes presented in a given stimulus array. The large quantity of shapes in Experiment 1 was intended to make context-checking more difficult for directors, and pressure them to reduce cognitive overhead by adopting conventions. This experiment, however, examines the performance of addressees, and it was useful for purposes of analysis to simplify the materials, such that each shape would have an equal probability of being subject to a fixation prior to the presentation of the audio stimulus. The number of shapes per item was reduced to four, and they were spaced equally from a central fixation point in the middle of the screen. The shapes were displayed at the four corners of an imaginary square, measuring approximately 16.3 by 16.3 cm. There was no grid bounding the regions of the square, as in Experiment 1. An example of the visual array is shown in Figure 3.1.
3.2.2 Design

Like the previous experiment, this experiment was a simple two-factor within-subjects design (see Fig. 2.1 from Chapter Two). The two independent measures were reference turn and community membership with the communicator. In the acquisition phase, items were organized in blocks. The acquisition phase allowed addressees to adapt their comprehension strategies to the conventions established by the first speaker. As in Experiment 1, the acquisition phase was followed by a transfer phase, whose purpose was to examine whether the acquired comprehension strategies carried over to a new speaker who would lack mutual knowledge of the conventions.
3.2.3 Procedure

3.2.3.1 Acquisition Phase

Upon arriving at the lab, participants were seated approximately 0.5m in front of a computer display. With their right hands, they controlled a computer mouse. Once they were comfortably seated, the experimenter gave them instructions on how to perform the task. They were informed that the purpose of the experiment was to investigate how people identify shapes based on another person’s descriptions, and that they would be listening to the pre-recorded descriptions of two previous participants in the experiment. They were led to believe that the descriptions were from actual participants, while in reality, the descriptions were read by confederates (see section on Materials). To avoid the possibility of the participants believing that they would be listening to recordings that were patched together by the experimenter, they were told that the previous participant described the same objects that they would see, and that they would hear these recordings in their entirety (see section 3.2.4.1 for discussion of further techniques used to enhance the inherent temporality of the descriptions). Their task was to listen to the descriptions and decide which of the four shapes was the target shape.

They were told that there would be a 30-second break at some point during the experiment in order to load the sound files for the second speaker.

The eyetracking equipment was then explained without drawing attention to the recording of eye movements. Participants were allowed to ask any questions before they signed consent forms. Finally, the eyetracking equipment was mounted and the
experiment began. The first trial of the acquisition phase was a filler, to prevent data loss in case the participant encountered any initial problems with the task.

Each trial began with a blank screen. A small, red fixation circle flashed in the middle of the screen. Participants were told to center the mouse over this point. This also had the consequence of centering their eye at the beginning of the trial.

Simultaneous with the presentation of the visual stimulus, the eyetracker began recording eye position in a digital format. Next, there was a 125 ms delay, and then playback of the audio began. The participant selected a target object with the mouse. To indicate that the selection had registered, the shape briefly flashed several times, and then the screen was cleared to begin a new trial.

After collecting a substantial amount of data, it was observed that many addressees adopted the strategy of remaining fixated on the center point and using peripheral vision to locate the target shapes. This resulted in substantial data loss. A subtle modification was introduced in the procedure for the last few subjects as an attempt to remedy this problem. The software was altered so that the fixation point would disappear once the shapes appeared, in the expectation that this would decrease the likelihood of using peripheral vision.

After completing all of the blocks of the acquisition phase, a message box appeared on the computer screen that invited the participant to relax and take a thirty-

---

21 See section 3.2.6 for further details.
second break while the computer loaded the second half of the experiment. The box had a timer that counted down the seconds until the beginning of the acquisition phase.

The acquisition phase had eight blocks of 17 trials each, for a total of 136 trials, requiring approximately 12-15 minutes to complete.

3.2.3.2 Transfer Phase

The transfer phase began immediately following the 30-second break. The purpose of this phase was to test whether the use of the conventions established in the first part of the experiment carried over when the addressee interpreted descriptions from a new speaker.

Immediately before the first trial of the transfer phase began, as the final few seconds ticked off of the 30-second timer, subjects heard a sound file in which the experimenter briefly addressed the new speaker. This snippet of conversation was intended to give participants a chance to hear the new speaker’s voice before performing any transfer trials. Furthermore, this (extremely banal) exchange drove home the message to addressees that they would now hear descriptions from a new speaker who was naive to the conventions from the first part of the experiment. The interchange went as follows: Experimenter: “Okay, so you’ve read the instructions?” Confederate: “Mmhm.” Experimenter: “Okay, you can start whenever you’re ready.” Confederate: “Okay, I guess I’m ready.”
After completing the transfer trials, the experimenter debriefed the participants, told them the purpose of the study, and showed them some of the data that they had contributed. Participants were allowed to ask questions about the experiment. The entire procedure lasted under 30 minutes.

3.2.4 Materials

In most respects, the kinds of shapes used in this experiment were identical to the previous experiment. One difference was that the nominal shapes were no longer used as experimental items, because eye fixations for nominal trials would not yield a clear interpretation that was relevant to the hypotheses of the experiment. Instead, they were used as fillers. Another difference was that each trial had exactly four shapes.

Unlike the previous experiment, the scalar exemplars were not staggered in the “size-shade” space from which they were drawn (Figure 2.3). Instead, a small square was the same size as a small circle, and a dark square was the same shade as a dark circle. It was important to equate the probabilities of looking at the different kinds of scalar shapes, because some of the measures were based on the assumption that certain shapes were equally good candidates for scalar terms. If the smallest circle were to be smaller than the smallest square, then this would introduce a bias to consider the smaller circle upon hearing the word “small”. Beyond these differences, the shapes used in this experiment were identical to those in the first experiment. The rest of this section focuses mainly on the composition of the audio recordings.
One male and one female speaker, who were both adult native speakers of American English, read the scripts. Their descriptions were recorded to sound files. They read from scripts that were constructed by the experimenter.\textsuperscript{22} The scripts were modeled after the descriptions recorded in Experiment 1, both in terms of content and form. Like the naturally produced descriptions, they included hedges, pauses, and repairs. Because the speakers were reading from a script instead of actually describing shapes, care was taken to ensure that they used the correct prosodic contour for the different items.

The descriptions were stored in separate sound files. The sound files were edited such that voice onset occurred at the beginning of the file. The interstimulus interval between presentation of the visual array and the start of the sound file was set at a constant value of 125ms. This was chosen so that the experiment could proceed quickly. While this was quite faster than the average VOT in the first experiment, it did not appear to detract from the naturalness of the stimuli. Because the ISI is set to a constant value, any differences in fixation latency among the probe and transfer items will be due to the manipulation of the independent variables.

\textsuperscript{22} Should the reader object to the “artificiality” of stimuli that were composed by the experimenter, it should be recalled that the experimenter, by then, had become an expert on referential descriptions, having transcribed and coded by hand approximately 1100 of the spontaneous descriptions that speakers from the first experiment produced. The descriptions in this experiment were modeled after these descriptions, when not simply plagiarized.
3.2.4.1 Acquisition Phase

There were three distinct kinds of trials used in the acquisition phase: training, probe, and filler trials. Examples of the three kinds of trials are shown in Figure 3.4. At the top the three different visual displays are presented, paired with a sample utterance immediately below. For expository convenience, the target shape appears in the upper-left corner. However, in the experiment the target shapes appeared randomly in one of the four spaces across all items.

Training trials. Training trials allowed addressees to learn about the stable and variable aspects of context associated with each category of target shape. As previously, each shape category was paired with a contrast shape of a particular kind. Two versions of the materials were constructed. In Version A, target circles always appeared with a contrast circle of a different size, and target squares with a contrast square of a different shade. Version B reversed these pairings. All training trials also contained two filler shapes. In 24 of the 80 training trials, one of these filler shapes was a scalar shape of the non-target category (each of the eight exemplars appeared three times), and the other was chosen randomly from the set of five filler objects. In the other 56 items, two distinct non-target shapes were randomly chosen from the set of filler shapes to appear in the item.\(^{23}\) A random process assigned the four shapes to their positions in the array. Each of the four exemplars from each shape category appeared as target an equal number of times throughout the training items.

\(^{23}\) It was later discovered that, inadvertently, three of the training items had two filler shapes of the same category. However, these training items constituted only 4% of all training trials.
An example of a training item appears in Figure 3.4a. In the figure, a small circle appears as target, in the context of a larger circle and two filler shapes. The addressee would hear this circle referred to as “the small circle”.

![Diagram of training, probe, and filler items]

**Figure 3.4. Sample Training, Probe, and Filler Items from Experiment 2.**

Each of the two speakers recorded descriptions for both versions of the materials. For each version, there were four main descriptions: “small X”, “large X”, “light Y”, and “dark Y”. (For Version A, replace X with “circle” and Y with “square”, and for B, replace X with “square” and Y with “circle”.) Thus, the speakers recorded eight different kinds of descriptions in all, four for each version of the materials. The experimenter recorded four takes of each kind of description, so that addressees would hear what appeared to be natural variation in the speakers’ descriptions. During the experiment, the software randomly chose one of these four versions during the presentation of a given training item. (Additionally, two takes of the speaker saying “the circle” and “the square” were recorded for the probe trials.)
**Probe trials.** The probe trials were designed to explore the degree to which addressees conventionalized their strategies for interpreting speakers’ descriptions. Specifically, these items were paired with descriptions that were intended to measure the degree to which addressees used physically co-present information (local context) and conventions (prior context) in interpreting speakers’ descriptions.

The visual stimuli for probe trials consisted a scalar target from one of the two shape categories, which appeared in the context of two contrasting non-target scalars of the other category, along with a fourth, “baseline” filler shape. Consider the item presented as Figure 3.4b. In this item, a circle that has been conventionalized in the training items as “the small circle” appears as target. For the first two blocks, addressees would hear the speaker refer to this shape as “the circle”, because it is the only circle in the display. This was important, as use of the conventionalized description before it was appropriate could raise questions about the “cooperativeness” of the speaker.\(^{24}\) However, like the speakers in Experiment 1, it is assumed that speakers would begin to refer to this circle as “the small circle” even when there is no contrast circle. After block 3, addressees begin hearing the description “small circle” to refer to the target. Because the scalar term “small” would select this shape only through convention, this shape is henceforth called the *convention-relative shape*.

\(^{24}\) The use of the simple NP descriptions made it impossible to establish true baseline values for fixations on the different shapes in the array, because the earliest trials became incommensurable with later acquisition and early transfer trials. For this reason, the data are analyzed differently from Experiment 1.
One of the goals of the experiment was to compare the degree to which addressees use conventions versus local contextual information when they interpret a speakers’ descriptions. For this reason, each probe item also included a “contrast pair”, consisting of two scalars of the non-target category. In this way, it was possible to measure the degree to which addressees interpreted “small” as picking out the smaller of two alternatives in the display, versus conventionally picking out a referent. Following Tanenhaus et al. (1996), addressees who maximize their use of local context should be more likely to make eye-movements to the smaller of the two squares in the display than to the conventional shape. The shape that is picked out by local information, such as the small square in Figure 3.4b, is termed the local-relative shape, while the shape that provides the contrast (e.g., the larger square) is called the contrast shape. The filler shape in each probe item establishes an important baseline, and thus is called the baseline shape.

It is important to note that when the addressee hears the phrase “the small circle”, it is only possible to compute the relative use of conventions and local information prior to hearing the noun “circle”. Once the addressee hears “circle”, any measure comparing the use of conventional or local information would be invalid, because “circle” unambiguously selects a target. The point in the description at which the noun begins is called the point of disambiguation.

The point of disambiguation for all of the scalar descriptions in the experiment was determined by hand, using software specially designed for this purpose. The software made it possible to select a “window” of data from the sound file, listen to it,
and adjust the window until only the first two words and initial sibant of the noun were audible. Because saccades are ballistic movements of the eye which require about approximately 200 ms to plan (Matin, Shao, & Boff, 1993), it is possible to increase the amount of data by adding this value to the point of disambiguation. To be conservative, a value of 180 ms was added to the point of disambiguation. This sum of these two values is called the *cutoff point*. Any fixations before the cutoff point can be safely assumed to have been planned before hearing disambiguating information.

**Filler trials.** A sample filler items is shown in Figure 3.4c. Though the two odd-looking nominal shapes from Experiment 1 were demoted to filler shapes in this experiment, they still retained an important role, along with the three other filler shapes. Principally, filler shapes served three critical goals: (1) they sought to convince addressees of the preservation of the original time-sequence in which the descriptions were supposedly recorded; (2) they reinforced the understanding at the beginning of the transfer phase that the transfer speaker was not be privy to the conventions established by the acquisition speaker; and (3) they made it difficult for participants to adopt the strategy of exclusively paying attention to scalar shapes.

This first goal was met by recording up to six different descriptions of the nominal and filler shapes, each shorter in length than the last. See Table 3.1 for the complete script for filler items. In the table, a picture of the shape appears in the top row, followed by the ordered descriptions from the acquisition phase and transfer phase. The descriptions were modeled after speakers’ naturally produced descriptions from Experiment 1. For the two nominal shapes, the first two descriptions were highly
elaborate, and like the spontaneously produced descriptions, they contained “um”s and “uh”s, pauses, and qualifications. Each reference turn, the description diminished in length and became more and more codified. Some of the early descriptions ended with rising intonation, indicative of a “try marker” (Sachs & Schegloff, 1979). These descriptions are marked with a small cross in Table 3.1.

| 1 | “uh... this one looks like three horizontal rectangles stacked on top of each other, sort of... unevenly” | “it looks like a... circle, with wedges cut out of it, at the top and bottom” | “uh... looks like a plus” | “it’s like a... a horizontal bar” | “it’s a triangle” |
| 2 | “the three rectangles stacked unevenly” | “the circle with the wedges cut out” | “the plus sign again” | “the horizontal bar again” | “the triangle” |
| 3 | “the three rectangles stacked unevenly” | “the circle with the wedges cut out” | “the plus sign” | “the horizontal bar” | “triangle” |
| 4 | “the three rectangle shape” | “circle with the wedges” | “plus sign” | “horizontal bar” | “triangle” |
| 5 | “three rectangles” | | “plus sign” | “horizontal bar” | |
| 6 | | | | | |

**Table 3.1. Script for Descriptions of Nominal Filler Shapes**

Each additional time that a filler shape appeared as target, the next description in the list was chosen until reaching the bottom of the list, which was used thereafter.
The nominal shapes appeared as targets at the beginning of each block, randomly interleaved with the two probe trials.

There were a total of 40 filler trials in the acquisition phase (16 nominal trials and 24 fillers).

3.2.4.2 Transfer Phase

There were 16 transfer items in the experiment. Ten of these were experimental items falling into three different categories depending on the identity of the target shape: extreme target, local target, and conventional target. Only the local and conventional target items were analyzed in this experiment, but descriptions of the extreme target items are included for the sake of completeness. Of the 16 transfer items, six were filler items. Examples of the transfer items are displayed in Figure 3.5.

Figure 3.5. Sample Items from the Transfer Phase, Experiment Two.
Extreme target. During the acquisition phase, it is assumed that addressees “absolutize” the meanings of relative scalars such as “small” or “dark”, assigning them to particular values along a continuous dimension. Through this process, scalar exemplars bearing particular values are conventionally associated with corresponding scalar labels. To give an example, the label “small” would become associated with a circle of a particular size, rather than to the smaller of two alternatives in the visual array. Thus, according to the mutual knowledge hypothesis, these conventionalizations would constitute “conceptual pacts” between speaker and addressee. According to the theory proposed here, these would simply represented entrenched comprehension strategies, which should show no partner-specificity.

One way to distinguish these alternatives is through the use of new scalar shapes with extreme values along the conventionalized dimension. Figure 3.5a provides an example. In the figure, the circle that was conventionalized as the “small circle” is paired with an even smaller circle (bottom right). When the addressees hear the description “small circle” by the transfer speaker, the mutual knowledge hypothesis predicts that they should unambiguously select the smaller circle as a referent. However, if they simply interpret “small” as referring to the conventionally small circle, then they would temporarily perceive the utterance as ambiguous.

Two extreme target items were constructed. The first was identical to Figure 3.5a. The second presented the conventionally “dark” square accompanied by an even darker square. The size relation was the first non-filler item in the target phase, and the
other shade item appeared after the first four local target shapes, before the unique target shapes.\textsuperscript{25}

\textbf{Local target.} The shapes used in the four local target items were identical to the probe trials from the acquisition phase (see Figure 3.5b). However, the recordings that were paired with the visual stimuli differed in an important way.

For purposes of analysis, it would be most convenient to make the transfer trials as comparable to the probe trials of the acquisition phase. However, to do so would give rise to a problem. If the transfer phase speaker, like the acquisition phase speaker, were to use the convention before it was possible for her to have acquired it, this would constitute a violation of the Maxim of Quantity and could raise questions as to whether she was a cooperative speaker. The concern was that if addressees inferred that the transfer speaker was uncooperative or in league with the experimenter, this would invalidate the findings.

To circumvent this problem, the target was selected as the smaller of the two shapes in the local contrast set, instead of the lone scalar. Thus, instead of saying “the small circle” to describe Figure 3.2b, the speaker would say “small square”. Thus, these items are only comparable with the acquisition probe items when only eye-movements

\textsuperscript{25} As in Experiment 1, the extreme trials are qualitatively different from the probe and transfer trials, because selection of the conventional target would lead to an error. For this reason, they were analyzed separately, though the analysis is not reported here.
that occur before the cutoff point are considered. The main question of interest is whether, before the cutoff point, the participant is more likely to fixate on the conventional target (the circle in Figure 3.2b) than the local target (the smaller of the two squares).

The restricted search hypothesis predicts that before the cutoff point, addressees will be more likely to fixate on the local-relative shape than the convention-relative shape, because they cannot assume that the new transfer phase speaker would be aware of the convention. A truly cooperative speaker would not refer to the circle as “small” because such a description would violate the Maxim of Quantity. Therefore, if addressees assume that the new speakers are cooperative, then they should expect them to use “small” to pick out the smaller of the two alternatives presented in the display.

In contrast, the theory of dynamic coordination would predict that assumptions of cooperativeness and mutual knowledge would have no effect on early comprehension processes. If it is the case that during the acquisition phase, addressees used conventions in referential search, then they should show no greater preference for the local-relative shape than the convention-relative shape. This is because they would simply map the descriptions from the new speaker onto old comprehension strategies. The comprehension strategy, however, should change over the course of the transfer phase, because using the old strategy would not efficiently yield a referent. This is

---

26 For purposes of precision, for both probe and transfer items the point of disambiguation is actually located after the initial silbant of the noun, because “circle” and “square” both begin with the same sound.
because early processes would select the conventionally small shape as the target, while
the local-relative shape is actually the target for these items. This revision would result in
an increased use of cognitive resources.

There were four local target items. Two circles and two squares appeared as
targets.

**Unique target.** Though in the beginning of the transfer phase it was important
that the speaker avoid using the conventions, this was not necessary by the end of the
transfer phase. In order to collect more data and examine how addressees dynamically
update their comprehension strategies, four items identical to the local target items were
included at the very end of the transfer phase. In these, however, the speaker used the
convention to refer to the scalar that appeared uniquely. Thus, for the item in Figure
3.5c, the addressee would hear the description, “light square”. Once again, these items
are most informative when we consider only those fixations that occur before the cutoff
point.

The four unique target items appeared in a random order at the end of the
transfer phase, interspersed with fillers.

**Filler items.** As Table 3.1 makes clear, the two speakers were given very
distinctive descriptions for the different filler shapes. This was intended to reinforce the
belief that there could not have been any mutual knowledge between them. To be
extremely sure that participants knew this at the beginning of the transfer phase, before
they viewed any experimental items they saw two filler trials with the two nominal
shapes as targets. The fact that they heard new, elaborate descriptions for shapes that the
previous speaker had described differently made it implausible that the acquisition and transfer phase speakers would somehow have mutual knowledge of the conventions.

In addition to these two filler trials, there were four other fillers interspersed randomly throughout the transfer phase, for a total of six filler trials in all.

3.2.5 **Apparatus**

The apparatus was essentially the same as in the previous experiment. In this case, two sources of data were considered: addressees’ eye movements and reaction time. The experiment control software was modified so that it would synchronize the presentation of visual stimuli with the recording of eyetracking data segments. Once the stimuli appeared on the screen, there was a 125 ms delay followed immediately by playback of the audio stimulus. When the addressee clicked upon the target shape, data recording was terminated immediately and the screen cleared for a new trial.

3.2.6 **Participants**

The participants in this experiment were 40 male and female undergraduate and graduate students from the University of Chicago who participated for payment. They were recruited through the posting of announcements on the internet and on flyers around campus. After much data had been collected, it was noted that many participants used peripheral vision throughout a large portion of the task. That is, they selected targets with the mouse while remaining fixated upon the central fixation point. Because in these cases they made no eye movements, 11 participants were removed, based on the
criterion of having used peripheral vision on at least three trials. This left 29 participants (13 male and 16 female).

3.2.7 Analysis

In order to obtain converging evidence for the different hypotheses, three dependent measures were considered for the experimental items in the acquisition and transfer phases. All of the measures used eye-movement data to test hypotheses. Two of these measures were related principally to the amount of local contextual information that addressees consider before selecting a referent. These two measures were number of fixations and total fixation time. The third measure, fixation probability, examined the probability of fixating on different shapes before the cutoff point.

Fixation measures were computed automatically using software designed for this purpose. Four different “areas of interest” (AOIs) were defined as regions bounding the four positions of shapes. The beginning of a fixation was located as the first of six consecutive frames of eye data (100 ms) within a given AOI (discounting blinks). A fixation ended once the eye position was outside of the AOI for six consecutive frames.

3.2.7.1 Measures relating to use of local context

Number of fixations on non-target objects. When addressees develop expectations about the kinds of objects that are referred to as “small”, “large”, “light”, or “dark”, they can be said to have established a kind of convention, which would facilitate referential search. Consider the first instance that an addressee hears “small circle”. Because “small” is a relative term, the addressee would have to exhaustively check
context to be sure to select the smallest of all circles. As prior context accumulates, the addressee can terminate the search with some confidence upon locating a circle of a particular size, because the addressee would know that that size of circle is the smallest of the exemplars of that category. If the speaker had really intended to refer to the smaller circle, in order to be cooperative the speaker would need to identify an even smaller circle using other means (the comparative “smaller” or “smallest”, for instance). Thus, as addressees conventionalize the scalar terms, the amount that they look at non-target shapes should drop.

This measure can only be used to test hypotheses for the acquisition phase. A change in the number of fixations in the transfer phase is indeterminate with respect to the two hypotheses, because the probe and local target items select different targets. Though a larger number of fixations on non-target objects would be consistent with the mutual knowledge hypothesis, it is also consistent with dynamic coordination, inasmuch as addressees the description itself forces addressees to use local information instead of prior context. It would be possible to consider only fixations before the cutoff point, but this would create an artificial ceiling that would mask any possible effect.

Though a comparison between the end of acquisition and the beginning of the transfer phase does not discriminate between the two theories, it can still perform an important check. Should the results of the probability analysis favor dynamic coordination, there should be a concomitant rise in measures related to the use of local context from acquisition to transfer. To see why, consider the fact that when addressees hear the conventionalized scalars in the beginning of the transfer phase, they will be
looking for the corresponding conventionalized target (e.g., for “small”, they would look for the small circle). However, when they hear the disambiguating noun (e.g., “square”), the new information would clash with the shape that they were considering, and they would have to adjust their comprehension strategy.

Temporary ‘errors’ such as these should be followed by a shift in comprehension strategy over the transfer phase that would shift the weighting of the scalar terms from away from conventional information. Because each of the two scalars is used four times to select a local referent, and four times to select a conventional referent, the information value of these words is effectively zero. This should be reflected in an equal probability of looking at the shapes in the display before cutoff.

Total fixation time. The total fixation time for a given trial is simply the sum of all of the durations, in milliseconds, of all fixations on all shapes. Like number of fixations, the total fixation time measure also is useful for testing hypotheses about the use of local context, but is of limited use in discriminating between the two theories. Once addressees become familiar with the different scalar exemplars and the contrast pairings, they can use this information to guide their visual search for the target. Instead of having to examine all of the shapes in the visual display, they could simply look for a particular shape, and terminate this search once they locate the shape. They would need to make fewer relative comparisons, and this should be reflected in total fixation time.
3.2.7.2 Fixation Probability

As described above, a disambiguation point was determined for each description used in the experimental items. The cutoff point was computed by adding 180 ms to the disambiguation point, accounting for “saccadic overhead” (Matin, Shao, & Boff, 1993). Thus, any fixations that began before this point would reflect information relating only to the scalar adjective and not to the noun, which uniquely identified the target.

Fixation probability represents the probability of fixating on a given shape before the cutoff point. It is computed for each shape in a given trial by calculating the percentage of digital frames before the cutoff point wherein the addressee fixated on the shape. Thus, if the addressee fixated on the filler shape for half of the digital frames of the pre-cutoff period, this shape would be assigned a fixation probability of 0.5. As it is impossible to fixate on two shapes at once, the probability values are mutually exclusive. However, they do not necessarily sum to 1, because addressees spend much of the pre-cutoff period fixating on the center fixation point.

Unlike the previous two measures, this measure gives powerful insight into the earliest moments of comprehension. Most importantly, it illuminates how addressees use contextual information when comprehending speakers’ descriptions.

When a speaker refers to a shape as “small”, addressees can use two different kinds of information to decide which referent the speaker intends. On the one hand, they could use the information in the visual display, or local context. On the other, they could
use their knowledge of the correlation of “small” with “circle”, which would make it necessary only to find a circle of a particular size. Because addressees would have a particular size in mind, it would not be necessary for them to exhaustively check the context to make sure the selected the smallest circle. They would therefore minimize effort in using the convention. Over the course of the acquisition phase, addressees should show heightened use of prior context.

When addressees face a new speaker in the transfer phase who refers to a shape as “small”, they also have these two kinds of information at their disposal. However, only one of these sources is mutually known. Because local contextual information is physically co-present with the new speaker, they can infer that it is mutually known. In fact, other than the first two filler trials (neither of which used scalars) and one of the two extreme trials27, this is the only non-trivial information that the addressee could infer as “mutually known”.28 If addressees use this information exclusively—that is, if they engage in restricted search—they provide evidence that they assume their speakers are following the principle of optimal design. In other words, they would show that addressees expect speakers to design their utterances with respect to information that is mutually known, and only information that is mutually known.

27 It is possible that the initial two extreme trials could somewhat dislodge the conventionalization of the scalar, because the targets of these scalars were more extreme-valued than the conventional shape. If anything, these trials would enhance addressees’ use of local information, working against dynamic coordination and in favor of mutual knowledge.

28 Trivial information would include things like the fact that they are both members of the community of English speakers, they are both participating in a psychology experiment, etc.
On the other hand, they can use the conventions they established with the speaker of the acquisition phase. This would constitute evidence that they do not expect speakers to tailor their utterances with respect to mutual knowledge. Rather, it would indicate the startling and anomalous result that addressees experience an “illusory transparency of convention”—they would expect speakers to optimally design their descriptions with respect to information which only they know. While this result is surprising with respect to normative models of language use, it is predicted by the theory of dynamic coordination.

The degree to which addressees consider either source of information would be manifest in the pre-cutoff fixation probabilities. Consider the example in 3.5b, where the scalar “small” has been conventionally associated with the small circle. To the degree to which addressees consider referents conventionally associated with the scalar, there should be a rise in the fixation probability on the circle (the convention-relative shape). To the degree to which addressees consider referents on the basis of the local, physically co-present information, there should be a rise in the fixation probability on the small square (the local-relative shape).

However, the probability measure is best interpreted with respect to a chance baseline, because addressees’ eyes tend to survey the objects, perhaps even before they have any candidate shape in mind as a referent. This baseline provided by the filler shape, which is the triangle in Figure 3.5b. To the extent to which the fixation probabilities for the conventionally “small” circle or the relatively “small” square are higher than this baseline, it indicates that they are being considered as referents. A
“relative preference” score for the conventional- and local-relative shapes was computed by subtracting this baseline probability from their probability scores.

Assuming the perspective adjustment hypothesis is correct, the fixation probability measure allows further investigation of the time course of adaptation of comprehension strategies and the role of communicative (or interpretive) success in calibrating these strategies. By examining how these pre-cutoff fixations change over the course of the transfer phase, it is possible to examine how quickly comprehension strategies are dynamically updated to reflect changes in the communicative environment. In the first few trials of the transfer phase, the size-shape pairings are reversed, and speakers tailor their descriptions to purely local information. If addressees adjust to this quickly, then there should be some decrease in the number of fixations on conventional objects by the second two of the four local target trials. The last four unique target trials use the conventions once again. This should also reveal the dynamic updating of comprehension strategies.

3.3 RESULTS AND DISCUSSION

Unlike Experiment 1, in this experiment preliminary analyses revealed differences between Transfer blocks 1 and 2. Because of this, it is would be unwise to group the first two blocks of the transfer phase together. Therefore, the blocks were not grouped together as in Experiment 1, but are analyzed separately. For the same reasons discussed in Experiment 1, the results are reported with participants as a random factor, but not items.
3.3.1 Local Context

It was hypothesized that addressees dynamically calibrate their comprehension strategies to speakers’ use of certain words. One of the ways they might do this is to establish conventions that assign particular values to scalar terms. If this is the case, then there should be a trend over the course of the acquisition phase for addressees to make fewer fixations before selecting a target. They would also spend less time fixating on shapes. On the other hand, if addressees do not use conventions in interpretation, then when communicators used these conventions the amount of time fixating on shapes should not drop, because the spoken conventions would be “redundant” information.

Figure 3.6. Overall Number of Fixations on All Shapes in Acquisition Phase
Figure 3.6 shows that the number of fixations on shapes before decision decreased from the beginning to the end of the acquisition phase. The number of fixations declined from an average of 2.95 fixations in block 3 to 2.48 fixations in block 8, just prior to the transfer phase. The decreasing linear trend was statistically significant, $F(1, 28) = 4.55, MSe = 0.976, p < 0.05$. Consistent with both models, there was also a significant increase from 2.48 to 2.93 from Acquisition 8 to Transfer 1 ($t(28) = 2.33, p < 0.05$).

Likewise, the total amount of time spent looking at non-target objects decreased from an average of 1751 ms in Acquisition 3 to 1654 ms in Acquisition 8 (see Figure 3.7). However, the linear trend analysis failed to reveal a significant effect $F(1, 28) = 2.84, MSe = 210847, p > 0.10$, though the decrease of 187 ms between block 3 and block 8 was marginally significant ($t(28) = 1.39, p < 0.10$). Surprisingly, there was no increase in total fixation time from Acquisition 8 to Transfer 1; in fact, the trend was in the opposite direction.
Figure 3.7. Total Fixation Time on All Shapes, Acquisition Phase.

The significant decrease in number of fixations provides support for the idea that addressees used conventions in understanding descriptions. Even though they clearly fixated on a fewer number of shapes, evidence that they spent less time fixating on objects barely approached significance. It is not clear why the drop in number of fixations was not accompanied by a significant parallel trend in fixation time. It might suggest that although addressees looked at fewer shapes, the fixations on these shapes became slightly longer over the course of the acquisition phase.

The number of fixations was observed to rise back up to its baseline (block 3) value when the addressee interpreted a new speaker’s descriptions. This finding would be subject to two different interpretations by the two theories. Dynamic coordination would interpret it as temporary confusion, because the scalar was used to refer to a non-
conventionalized referent. Mutual knowledge would view it as an increase in the use of physically co-present information with the new speaker, because the addressee would assume that the utterance was designed with respect to the mutually known, physically copresent information in the local context.

The fact that the number of fixations decreased during the acquisition phase demonstrates that addressees were using less and less information from local context. However, it is possible that addressees were simply using peripheral vision more, instead of foveating directly on the objects. The following analysis seeks to show that they were fixating less because they were using conventions based on the prior context.

3.3.2 Fixation Probability

Over the course of the acquisition phase, it is predicted that the probability of looking at the conventional target would increase relative to the baseline value of looking at the filler shape. This relative measure, referred to as a “preference score”, was obtained by subtracting the baseline from the probability of looking at the conventional shape for each trial, and then averaged. The preference scores are displayed in Figure 3.8.
Figure 3.8. Preferences for Fixating on Local- and Convention- Relative Shapes

Except for what appears to be an anomaly in block 7, from block 4 to block 8 addressees consistently showed a higher probability of looking at the convention-relative shape than the baseline shape. It is unclear what could explain the strange pattern in Block 7. The mean preference score over the acquisition phase for the convention-relative shape was 0.06. This score was significantly different from baseline, \( t(28) = 4.91, p < 0.05 \). The mean preference score for the local-relative shape was 0.05, \( t(28) = 4.12, p < 0.05 \). Neither score showed a significant linear trend (both \( F_s < 1 \)). Notably, the preference for the convention-relative shape was not significant in Block 3 (\( t(28) = 0.97, \text{n.s.} \)). Thus, it appears that addressees locked into an effective strategy by block 4, because the preference for the conventional shape did not significantly rise over the
course of the acquisition phase. Unfortunately, the experiment does not allow us to examine how that strategy may have evolved prior to block 3. Moreover, contrary to predictions, the preference for the local shape did not drop from block 3 to block 8. This unexpected result suggests that addressees only partially conventionalize the meanings of the scalars, but retain flexibility in their comprehension strategies. The question, then, is whether the preference for the conventional-relative shape disappears when the addressee faces a new speaker in the first block of the transfer phase.

In the first block of transfer, the probability of looking at the convention-relative shape was 0.06 points higher than baseline ($t(28) = 2.14, p < 0.05$). While this relative preference remained roughly the same from Acquisition 8 to Transfer 1, surprisingly, the probability of looking at the local-relative shape did not increase; in fact, it dropped to 0.03, which was not significantly different from baseline ($t(28) = 1.27, \text{n.s.}$).

The picture becomes clearer if the probabilities for fixating on the various shapes are considered over the entire interval. Figure 3.9 presents the fixation probabilities over the course of the first 1000 ms of the first two transfer items (the first transfer “block”). Time is plotted along the horizontal axis and fixation probability on the vertical axis. Although the figure is a composite of the first two local target items of the transfer phase, for expository purposes, it is described with reference to the item in Figure 3.5b, in which the “small square” is the target. This shape appears with a circle that was conventionally referred to as the “small circle”. The other shapes in the example are a large square (local-relative) and a filler shape (baseline).
Three important temporal events are shown as vertical lines crossing the body of the chart. These events are (voice) onset at 125 ms, the point of disambiguation at 557 ms, and the cutoff point, which is the point 180 ms after the point of disambiguation. Any fixations starting anytime before the cutoff point were planned before disambiguating information was available. Text corresponding to the description used in the example is displayed, banner-like, across the top of the chart. The probability of looking at the circle is shown as a dark, thick, solid line; the square is shown as a thick, gray, solid line; the large square, a thin, dashed line; and the filler shape as a thin, solid gray line.
The first thing to note is that between the onset and the cutoff point, the probability of looking at the small square is consistently lower than the probability of looking at the small circle. Thus, the convention appears to override local salience, as in Lewis’ (1968) formal definition. In fact, while the probability of looking at the conventionally small circle was significantly higher than baseline, the probability of looking at the small square was not. Only after the cutoff point does the probability of looking at the small square become systematically greater than that of the small circle. While the probability of the small square continues to increase after cutoff, the probability of the small circle begins to decrease.

This constitutes strong evidence that addressees did not adjust their comprehension strategies when they faced a new speaker, as the mutual knowledge hypothesis suggests that they should. Rather, they expected that the new speakers’ use of “small” would have the same conventional meaning as the previous speaker. In other words, they experienced an “illusory transparency of convention”. It would appear to support the prediction of dynamic coordination that addressees only adjust their comprehension strategies when these fail to minimize effort. This is supported by the finding that the number of fixations increases from Acquisition 8 to Transfer 1.
Figure 3.10. Dynamic Adjustment During Transfer.

As further evidence for dynamic partner adaptation, consider Figure 3.10, which represents how comprehension shifted over the course of the transfer phase to reflect the speakers’ differential use of local and prior context. The chart shows the overall pre-cutoff fixation probabilities for the four different shapes in Acquisition 8, Transfer 1, and Transfer 4. In Transfer blocks 1-2, the speaker uses the scalar to refer to the local target (local target items; see Figure 3.5b). However, the addressee is still more likely to look at the convention-relative shape than the local-relative shape. By the end of the transfer phase, the addressee is no more likely to look at any one shape than any other. It appears that addressees dynamically reweighted the conventional meaning of the scalar, eventually realizing that its information value was effectively nil—in half of the cases, the target was the local-relative shape, in the other half, the conventional shape.
Clearly, then, addressees will dynamically adapt to a communicator’s use of a term attempting to minimize their effort without compromising communicative success.

3.4 CONCLUSIONS

As with the production experiment, the results of this experiment support dynamic coordination over mutual knowledge. During the acquisition phase, addressees assigned conventional values to the scalar terms. They appeared to use their knowledge of prior context, in order to simply referential search. This knowledge of prior context was established through the gradual adjustment of a comprehension strategy that strongly weighted size-relative scalars with circles, and shade-relative scalars with squares. The use of the convention reduced the amount of local context that addressees considered, as reflected in the decrease in the number of fixations on shapes over the course of the acquisition phase. A similar trend for total fixation time was not statistically significant.

The finding that conventions appeared to minimize the use of local context in the acquisition phase, at least when number of fixations was taken into account, could also be explained by mutual knowledge. It would explain the greater efficiency of referential search through the establishment of mutually known conventions (or conceptual pacts) with the communicator. However, if this were the case, then addressees should have abandoned their use of the convention when they interpreted descriptions from the new communicator in the transfer phase. Clearly, they did not. As Figure 3.9 shows, addressees were most likely to fixate on the convention-relative shape
in the period before the cutoff, and only after the cutoff did they become more likely to fixate on the local-relative shape.

One possible interpretation of the results is that addressees assumed that the second communicator would have somehow known the conventions. Perhaps they inferred that this communicator had previously participated with a different addressee and therefore had experience with the shapes. Though the experiment cannot completely rule this out, an element of the procedure makes this interpretation of the results unlikely. The first two trials of the transfer phase were nominal shapes that were described in a way that clearly reflected the idea that the communicator had never seen these shapes before. The descriptions were long, and filled with hesitations and contained the pause markers “um” or “uh” (see Table 3.1). These descriptions ended with a “try marker” intonation, which would imply that the descriptions had not yet been conventionalized.

Another possible interpretation would be that maybe addressees inferred that the two communicators somehow had mutually established conventions with each other for describing the shapes in the experiment. This is also a very unlikely inference, for the same reason—the two communicators gave very different descriptions for the filler shapes. The extreme differences between these descriptions should have made it very clear to addressees that the inference of mutual knowledge between the two communicators was unwarranted.

Yet another possible objection could be that the use of recorded stimuli makes it difficult to generalize the results to natural interaction. This point, at least, must be conceded. It could be argued that perhaps the use of a real communicator would make
mutual knowledge more salient. However, in tasks using real communicators, similar effects have been found (Keysar, Barr, Balin, & Brauner, in press).

It is also important to keep in mind another way it might fail to generalize to natural situations. Dynamic coordination suggests that the multimodal feedback present in natural interactions allows interlocutors to be more egocentric—the collaborative situation allows interlocutors to diagnose and correct mistakes before they occur. However, even in the absence of feedback, addressees were still egocentric in how they interpreted descriptions. Thus, it is not clear that a more naturalistic setting would favor mutual knowledge over dynamic coordination.

In closing, mutual knowledge does not appear to be the “context for comprehension” for understanding conventions. Instead, like communicators, addressees are egocentric in their use of language. An examination of early processes of comprehension suggested that they exhibit an ‘illusory transparency of convention’. That is, they expected that a word would have a conventional meaning even when the person who used it was unlikely to know the convention. This is because the mechanisms behind dynamic coordination are not sensitive to the identity of the interlocutor, but simply use prior context in a way that makes referential search easier. Though addressees may initially experience difficulties understanding referential descriptions from new addressees, they retain enough flexibility in their comprehension strategies such that they can quickly adjust to the different information value of the descriptions.
CHAPTER FOUR

CONCLUSIONS

The experiments reported in the last two chapters were designed to meet the two criteria outlined in the Introduction: (1) they provide a rigorous control of task-relevant mutual knowledge between interlocutors; and (2) they deconfound the use of information because it is mutually known from its use due to perceptual or attentional salience (Keysar, 1997). The experiments met the first criterion by carefully controlling the knowledge accumulated by interlocutors in a referential communication task, such that knowledge of conventions between two interlocutors in the transfer phase could not be inferred. Each condition also provided a baseline such that use of the convention (e.g., scalar term) beyond the baseline level in transfer would provide evidence for the egocentric processing of information. The experiments met the second criterion of deconfounding personal salience with mutual knowledge. The use of a convention because it was mutually known versus because of its salience would result in different patterns of behavior in the transfer phase.

In this section, I review the findings from both experiments and discuss their generalizability to non-experimental settings of language use. I also attempt to delimit the scope of the theory of dynamic coordination. Then I discuss the implications that the findings have for theories of language coordination and convention use. Finally, I
indicate how the theory of dynamic coordination suggests avenues for future research. The chapter ends with a discussion of the larger implications of dynamic coordination theory for how theorists approach and explain conventions and related coordination phenomena in language use.

4.1 FINDINGS

4.1.1 Language Production

The first experiment tested various hypotheses regarding aspects of convention use in language production. These hypotheses concerned why (on what basis) communicators establish and use conventions; whether communicators use them in ways that are partner-specific; and how convention use is related to on-line speech production. Mutual knowledge and dynamic coordination theory made different predictions in each case.

For the shortening and conventionalization of descriptions, mutual knowledge predicted it would be achieved by a collaborative process of “grounding”, whereby each reference turn gives interlocutors the opportunity to renegotiate descriptions. What changes is the common ground between the partners. On the other hand, dynamic coordination predicts that what changes are the verbal plans themselves, in a way that is insensitive to mutual knowledge. Communicators recycle verbal plans, and apportion effort to refashion descriptions when they fail. The two theories made different predictions about the lengths of communicators’ descriptions when they described old referents to new addressees.
If it were the case that communicators followed the principle of optimal design, then their descriptions for the new addressee should have been as long as their first descriptions with the old addressee. However, the new descriptions were significantly shorter, suggesting they were recycling old plans instead. Furthermore, communicators’ voice onset time for descriptions with new addressees failed to increase significantly from the last reference turn with the previous addressee. In addition, though the initial descriptions for the new addressee were slightly longer than the last description with the previous addressee, further analysis suggested that at least 50% of this lengthening was driven by multimodal addressee feedback. In addition, the installment-like nature of communicators’ descriptions suggested that they were apportioning more effort to adjustment than to designing new utterance plans with respect to mutual knowledge. In sum, an analysis of nominal trials seemed to support dynamic coordination.

The scalar trials in Experiment 1 tested hypotheses regarding the source and partner-specificity of conventions. Consistent with dynamic coordination, communicators were observed to conventionalize their descriptions more and more over the course of the acquisition phase, suggesting that they were either establishing mutually known descriptions for referents, or maximizing their use of prior context. It was observed that when these same communicators faced new addressees, the rate of convention use initially remained the same, providing strong evidence against the claim that communicators initially designed their utterances according to the principle of
optimal design. Instead, that they used these conventions with addressees who did not know them suggests that they experienced an “illusory transparency of convention”.

The analysis left open the possibility that perhaps communicators were tracking mutual knowledge instead of maximizing prior context, but simply did not use this information when they faced new addressees. A linearly decreasing trend in VOT, which failed to increase in the transfer phase, supported the idea that they were merely recycling verbal plans and using prior context. The most informative measure in regard to the use of local context was an analysis of communicators’ eye movements prior to voice onset time. This analysis strongly indicated that during the acquisition phase, communicators were making less and less use of prior context when describing referents to the experimenter. The number of pre-VOT fixations decreased linearly during acquisition, and did not significantly increase once communicators faced new addressees. This unambiguously supports the idea that communicators used conventions egocentrically.

The number of post-VOT fixations showed an entirely different pattern, however. When communicators faced new addressees, the number of post-VOT fixations increased by approximately 38% (by 27%, when only the number of distinct shapes are counted). That post-VOT fixations showed a dramatic increase while pre-VOT fixations remained constant provides overwhelming evidence for the idea that communicators formulate descriptions egocentrically, and dynamically apportion effort to refashion these descriptions when prompted by multimodal addressee feedback.
In an analysis of the transfer phase fixations, communicators gradually increased the number of pre-VOT fixations as they had the opportunity to observe and diagnose problems that addressees encountered during referential search. The number of post-VOT fixations showed a parabolic trend, first increasing and then decreasing over the course of adjustment. It would be interesting in future experiments to try to correlate these two measures with the various kinds of feedback that addressees provide.

4.1.2 Language Comprehension

The experiment on comprehension also provided support for the predictions of dynamic coordination over mutual knowledge. In the acquisition phase, addressees appeared to conventionalize the meanings of scalar terms such as “small” or “light”. By the end of the acquisition phase, before the cutoff point they were more likely to look at the shape that was conventionally associated with a particular scalar than they were to look at a baseline shape. The convention for interpreting communicators’ descriptions facilitated referential search, because it allowed them to terminate search as soon as they found the shape with the conventionalized value. In other words, in cases where they heard “small...” and the first shape their eyes landed on was a circle that had been called “small” in the past, they would not need to look around to make sure that there was not an even smaller circle in the display. This facilitation of referent selection was reflected by a linear decrease in the number of fixations on shapes in the display.

When addressees interpreted descriptions from the transfer phase communicators, they experienced an illusory transparency of convention. Even though
the circle was the only circle in the display, upon hearing “small” they were more likely to look at it than the baseline shape. Interestingly, before the cutoff point they were no more likely to look at the shape that actually had a size contrast (the local shape, e.g., the small square) than they were at the baseline shape. This clearly refutes the idea that common ground is the context for interpreting conventions. Rather, addressees interpret conventions egocentrically, assigning them the conventional meanings that they have had most often in the past.

4.1.3 Implications, Generalizations, and Qualifications

The idea that communicators do not take mutual knowledge into account when they initially design their descriptions was first suggested by Brown and Dell (1987), and further corroborated by Horton and Keysar (1997). By forcing speakers to rapidly initiate their descriptions, Horton and Keysar demonstrated that preverbal plans were formulated egocentrically. These descriptions differed from unspeeded descriptions, which seemed to take mutual knowledge into account. Polichak and Gerrig (1998) criticized the generalizability of the results on several grounds, including that the speeded task was an imposition that compromised the generality of the findings (but see the response of Keysar & Horton, 1998). Polichak and Gerrig argued that it is not surprising that when communicators are placed under stringent time constraints that they should make certain pragmatic errors, and that to claim that these errors reveal the egocentrism of the initial plan is unjustified.
My experiment on language production shows an effect that is very similar to Horton and Keysar—in fact, the effect is more dramatic, because when communicators first encountered new addressees, they did not adjust their descriptions of scalar shapes at all. Yet this finding cannot be subject to Polichak and Gerrig’s criticism because the communicators in this experiment were under no time pressure to initiate their description. The only time pressure that some of them might have been under is the self-imposed one to finish the experiment, get the money, and go home. Of course, such an instrumental attitude toward conversation, even at its most extreme, it is one that exists in settings other than just the psycholinguistic laboratory. But even without the artificially imposed constraint of a time deadline—that is, even when communicators could produce utterances at their leisure—they still produced egocentric descriptions.

Furthermore, the results of this experiment show that the criticism that Gerrig, Ohaeri, and Brennan (in press) direct at Keysar’s original demonstration of illusory transparency (Keysar, 1994), has no empirical foundation. They argue that what Keysar’s results actually show is that addressees and overhearers are obsessive about preserving normative assumptions about their communicators. That is, they wish to preserve, at all costs, the assumptions that their communicators are cooperative and are following the principle of optimal design. This would mean that when a communicator seems to use information that violates that assumption, addressees and overhearers would take this as sufficient reason for assuming that that knowledge must have been mutually known in the first place. This is predicated on the assumption that real-life communicators, unlike those in Keysar’s experiment, do not violate mutual knowledge in
formulating descriptions. However, the communicators in this experiment, when left to their natural devices, behaved exactly like the protagonists in Keysar’s scenarios—they appeared to believe that their use of privileged knowledge (in this case, conventions) would be more transparent than they actually were. In short, they exhibited an illusory transparency of convention. Therefore, for an addressee to assume that communicators are always cooperative would be unwarranted, even perilous.

Addressees might still make these normative assumptions even though speakers appear to systematically disregard the principle of optimal design. If this is the case, then addressees should make many errors when they interpret speakers’ descriptions. Yet the transfer phase addressees in Experiment 1 had an error rate of around 5%—that is, they made virtually no errors. Though they had to interpret uncooperative descriptions such as “the small circle” when the target was, in fact, the only circle, they still chose targets instead of being simply paralyzed as they tried to find a context that would preserve the assumption that the addressee was optimally designing utterances. If, as the experiment suggests, addressees are routinely faced with egocentric communicators who use redundant descriptions, then we should expect them to be somewhat cautious and forgiving when dealing with violations of conversational maxims.

Is there any way that mutual knowledge explain the results of Experiment 1? Perhaps—one could reason, communicators only used the conventions because they assumed that the addressees knew them, even though they were absent when the conventions were acquired. Maybe the communicators suspected that the addressees
were actually confederates who were listening in the other room through a hidden earphone. It is unclear why any of them would make this assumption. The participants in this experiment were screened beforehand to make sure that none of them had participated in any other experiments conducted in our lab which used confederates. Some of them had never even been in a psychology experiment before. It is important to note that the standard theory could still stretch to explain these results, though it would require very unnatural and convoluted post hoc assumptions. These kinds of post hoc reconstructions of mutual knowledge bring the theory perilously close to the domain of unfalsifiability. The theory of dynamic coordination, for which these normative assumptions are hardly relevant, can explain these findings straightforwardly, without invoking such post hoc explanations about what participants may have believed about the nature of the experiment.

The generalizability of the second experiment, which looked at comprehension, is admittedly not as good as the first. This is because the need to experimentally control the speech stimulus made it necessary to use recorded stimuli instead of live communicators, eliminating the possibility for true collaborative interaction and online multimodal feedback. Moreover, in this experiment, it may have been more possible here than in Experiment 1 for addressees to assume that the communicators they were listening to were in league with the experimenter, such that the second communicator would have known the conventions. Nevertheless, measures were taken which make this kind of interpretation unlikely, such as the gradual shortening of descriptions for filler shapes. Still, it would be important to design a follow-up
experiment that would overcome these limitations. It is not immediately clear how to do
this experiment without compromising the relevant controls.

Another aspect of generalizability that should be addressed is the degree to
which the results from the two experiments would generalize to other kinds of
communicative situations. In written communication such as through electronic mail,
communicators and addressees do not have the possibility to interactively coordinate
reference. Perhaps in these settings, communicators and addressees are more cautious
about how they formulate and interpret utterances, because they know that they may have
only one chance to get it right. Depending on how aware interlocutors are of
egocentrism, it might be predicted that the degree of egocentrism interlocutors exhibit
would depend on the potential for multimodal feedback. When there is little possibility
for dynamic coordination, savvy communicators would apportion all of their effort in
designing custom-tailored plans. In these cases it might be the case that the theory of
mutual knowledge might provide a better account than dynamic coordination theory,
whose principles depend in large regard on the presence of multimodal feedback. These
issues need to be subject to empirical scrutiny.

The experiments suggest that mutual knowledge is not the essential condition
for conventions, or at least, for convention-like uses of definite descriptions. Dynamic
conversation theory appears to better explain the evolution and use of conventions, even
though conventions traditionally have been held up as one of the reasons why mutual
knowledge is essential in language use (Clark & Carlson, 1981).
Mutually known information, instead of guiding utterance formulation and comprehension, might be most important when coordination fails. In these cases, it may be necessary for interlocutors to compute what is mutually known in order to repair or replace erroneous utterances or interpretations. This suggests that the two theories may not be mutually exclusive, but instead may apply over different domains of language use.

Additionally, it may be the case that when interlocutors face huge asymmetries in expertise in some domain, such as when an expert is explaining a technical concept to a novice, they mutual knowledge (or lack thereof) might be extremely salient to the interlocutors. However, it is not clear that any apparent attentiveness to mutual knowledge would be a result of the kind of foresight assumed by the principle of optimal design. Instead, it is possible that while conversing, the usual multimodal indices of successful communication would not be present. As a consequence, the interlocutors might pay more attention to what is and what is not mutually known when formulating or interpreting utterances.

In short, the use of mutual knowledge may not be all-or-none, but could depend upon the nature of the specific coordination problem at hand. It would be important to establish more precisely the characteristics of coordination problems that interlocutors face so that the domains over which the different theories apply would be better understood.
4.2 FUTURE DIRECTIONS

While the two experiments examined communicators’ and addressees’ use of conventions, neither of them examined the back and forth exchange that typically occurs in conversation, where participants alternate roles. Thus, these studies do not fully address the phenomenon of lexical entrainment (Garrod & Anderson, 1986), in which the members of a dyad arrive at similar ways of describing a referent. It is not immediately clear that a convention that is acquired in production would be used in comprehension or vice versa. It would be important to show that the theory of dynamic coordination would extend to these cases.

Another important step in future research would be to provide further empirical validation for the various mechanisms underlying dynamic coordination for coordination problems other than definite reference. It is likely that dynamic weighting operates at levels other than just simple phrases, such as a communicator’s construction and reconstruction of a personal narrative, and perhaps at the level of discourses themselves, giving rise to codified conventional structures such as scripts (Schank & Abelson, 1977) that can greatly facilitate how interlocutors solve recurring problems of coordination.

The mechanism of dynamic information weighting has the necessary consequence of representational change. One of the more interesting predictions of the theory is that interaction changes how interlocutors represent entities. Another way of thinking of linguistic coordination is as a process of socially driven concept formation, in which interlocutors establish ad hoc categories (Barsalou, 1983, 1991) that support
coordination. This idea is not completely novel, as evidence for this has already been presented by Garrod and Doherty (1994) and Markman and Makin (1999). However, only the theory of dynamic coordination predicts that referential communication should result in conceptual change, because the only representations that change in mutual knowledge theory are representations about mutual knowledge itself. When interlocutors attempt to coordinate reference, they also attempt to coordinate how they represent the world, through dynamic weighting and maximization of prior context. This increases their success and minimizes the need to compute mutual knowledge. For the case of referential communication, the theory also predicts how representations should change—it predicts that the dimensions of a category should be weighted according to how successful they are in achieving reference. In the experiments, when circles appeared with circles of different sizes, the size of a circle was the only dimension of the category that had conventional value in establishing reference. The shade of a circle, on the other hand, was irrelevant. It would be fascinating to investigate whether these predicted representational changes are observed in non-linguistic categorization tasks.

As a general theory of coordination, the dynamic theory provides an alternative etiology of conventions to Lewis’ (1968) salience-precedence-convention sequence. Human interaction is intrinsically convention-laden, and thus it is an important priority for the cognitive sciences to develop a cognitively realistic theory of convention. One of the advantages of the dynamic theory is that the vague, cognitively complex, and potentially incalculable concepts of mutual knowledge and community membership do not play a central role in the theory. Instead, the theory takes advantage of the intrinsic
power of known cognitive mechanisms and the richness of multimodal environments. Because of this, the predictions of the model are easy to operationalize and test.

It is possible that the dynamic coordination approach could be extended to coordination problems in general. As in language, in non-linguistic domains of human interaction, conventions may not be precedents that are mutually known, but alternatives that have become highly salient and routinized through repeated use. Over repeated instances of a coordination, interlocutors dynamically weight alternatives based upon their success in the task at hand. They could use the success of these choices along with the rich feedback signals from other agents as a “teacher signal” to update the weights that they assign to different alternatives. Instead of precedents leading to mutually known conventions, precedents would become more salient alternatives each time they are used. The use of these precedents would become routinized and highly accessible in memory, such that agents would eventually experience an illusory transparency of convention. Thus, cognitive agents might be able to coordinate remarkably well by just doing what is salient to them—recycling and refashioning the successful portions of “action plans”—and perhaps only in rare cases, would they have to formulate a new plan tout court. In these cases, estimations of what other cognitive agents know, of what the others know that they themselves know, and so on, would be very important. Theories of coordination should not be based on these rare cases, however.

The theory of dynamic coordination was intentionally formulated in a manner that makes it amenable to computational modeling. This is an area that would be important to develop, for several reasons. First, it would help delimit the explanatory
scope of the model. A model which realistically embodies the assumptions of dynamic coordination would most certainly fail in interesting ways that would suggest improvements to the model (or alas, provide falsification). Most likely, the model would succeed in some ways and fail in others—illuminating the kinds of circumstances under which traditional mechanisms of coordination are either unnecessary or unavoidable. In cases where the theory does just as well as mutual knowledge in explaining coordination, it should be favored as more parsimonious because it reduces coordination to known psychological mechanisms, and does not needlessly proliferate additional intentionally-bound explanatory constructs.

Another reason computational modeling is important is that it would make it possible to draw out the implications of the theory at a larger scale, and trace how conventions spread throughout a community of cognitive agents who operate according to the very simple mechanisms of dynamic coordination. The consequences of the interaction of many such agents would be hard to predict and test without computational modeling. One could vary the structure of language communities, how spatially spread out the agents are, and their frequency of interaction to try to predict convention use on a massively social scale. These efforts could potentially bring the theory of dynamic coordination in line with general theories of language change and creolization.

4.3 CONCLUDING NOTE

We began this investigation with the question: What is the nature of the control mechanism underlying conversational coordination? We addressed this question
by examining how interlocutors establish and use conventions in producing and interpreting definite descriptions.

The theory of mutual knowledge—the dominant theory in research on language use—suggests that the common ground between interlocutors is the control mechanism which allows interlocutors to successfully coordinate reference. However, no evidence for optimal design nor restricted search was found in Experiments 1 and 2.

Clearly, then, interlocutors can get along rather well without these mechanisms. Instead, it seems strongly likely that the coordination that we observe in conversation is a phenomenon that ‘emerges’ from the operation of low-level cognitive and conversational mechanisms, the way that the group-like behavior of flocking birds can be explained by local interactions among birds. The power of this kind of analysis as a theory of language use has yet to be fully explored. The studies described in this thesis constitute a first step in this direction.

In closing, it should be acknowledged that the general insight—shared by Lewis, Schiffer, and Clark and his colleagues—that many problems of language use are fundamentally problems of coordination, is a brilliant one with far-reaching implications. My purpose here is not to deny its importance nor provide an account to replace it. Rather, I simply wish to show that the domain over which mutual knowledge might apply in language use might be vastly more limited than is commonly assumed, because the theory of coordination on which it was based has been misapplied.

When Schelling (1960) developed his theory regarding the structure of coordination games, he was aiming at the most general of possible theories. In the way
that we strip away the leaves of the artichoke to get to its heart, it was necessary for
Schelling to strip away layers of phenomena from problem domains in order to come to a
general understanding of how agents coordinate activity in an uncertain world.

Yet the kinds of coordination problems that language users face are
qualitatively different from Schelling tasks. They are rarely—perhaps never—one-shot,
irrevocable decisions made among a finite set of well-defined alternatives, with binary,
win-lose outcomes. Nor are language users allotted an indefinite amount of time and
mental energy to make their decisions—they must do so under the crush of time, most
likely while they are also concerned with other resource-consuming tasks. If problems of
linguistic coordination were more like Schelling tasks, then it would make a lot of sense
for language users to maximize the effort that they apportioned to computing mutual
knowledge when preparing verbal plans or comprehending utterances. That is, they
would behave as current models predict they should. The findings of Keysar and
colleagues, which suggest that they stray far from this mark, implies one of two mutually
exclusive possibilities: (1) that language users are neither optimally nor even adequately
designed for the task of coordinating language use; or (2) the structure of coordination
problems in linguistic coordination is actually quite different from how it has been
traditionally conceived.

Clearly, the latter alternative is the correct one. In some ways, the problems
we face as interlocutors are much harder than the ones that participants face in a
Schelling task. When we make linguistic decisions, we have an indefinite number of
alternatives to choose from, and a miniscule amount of time in which to make them. Yet
in other ways, our task is much easier than a Schelling task—the rich, multimodal contexts of language use enable us to gradually work toward coordination optima, instead of having to do it all in one shot. Because conversation is generally forgiving of mistakes, we can minimize our effort by apportioning it where it counts most, and dynamically weight alternatives in memory according to their success or failure. In this way, when we confront old problems such as having to refer to a referent over multiple turns, the scales are already tipped in favor of success. The solutions that worked best and most often in the past become the ones that we see first—perhaps, the only ones we see.
REFERENCES


