Convention and Cognition: Weak Definite Noun Phrases

by

Natalie M. Klein

Submitted in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Supervised by
Professor Michael K. Tanenhaus

Department of Brain and Cognitive Sciences
Arts, Sciences and Engineering
School of Arts and Sciences

University of Rochester
Rochester, New York

2011
Curriculum Vitae

The author was born in Lawrence, Kansas on June 22, 1982, but relocated to San Antonio, Texas as a small child, where she attended Saint Luke’s Episcopal and graduated from Saint Mary’s Hall. She was a student at Pomona College in Claremont, California from 2000 to 2004, during which time she spent a semester at National Chengchi University in Taipei, Taiwan. She graduated with a Bachelor of Arts degree cum laude in Linguistics and Cognitive Science and a minor in Chinese. She came to the University of Rochester in the Fall of 2004 and began graduate studies in the Department of Brain and Cognitive Sciences, where she focused on psycholinguistics. In May of 2008 she was awarded a Master of Arts in Brain and Cognitive Sciences. During her time in graduate school, she has taught undergraduates at both the University of Rochester and Syracuse University. Under the pseudonym Muffy Stopheles, she has also helped build a nonprofit women’s sports league, Roc City Roller Derby, where she served as Board Secretary in 2010 and Board President in 2011. Her research on human language processing has largely taken place with the support of her advisor, Michael Tanenhaus, as well as the support of other faculty and peers. A portion of her academic pursuits has culminated in the research presented here in her dissertation.
Acknowledgments

Nothing, especially research, happens in a vacuum, and I have been blessed to have a great deal of intellectual, institutional, and moral support throughout the course of this project. The University of Rochester, Pomona College, the National Institute of Health Training Grant T32 DC400035 and Sentence Processing Grant R01 HD027206 to Mike Tanenhaus, and the Sperry Fund have made a life of the mind possible for me.

I would especially like to acknowledge my colleagues on this particular project, Rachel Sussman and Whitney Gegg-Harrison, for providing the kind of collaboration only brilliant and thoughtful friends can provide. Ana Aguilar has also been an important sounding board for my work on weak definite noun phrases. Dana Subik has made the process of coordinating five studies stress-free with his endless cheer and organization. Greg Carlson has introduced me to many of my favorite mysteries about language and the mind and has offered me encouragement at several critical moments in my course of studies. My advisor, Mike Tanenhaus, will always be one of my favorite people to converse with, and his influence on my philosophy, intellect, and taste in scotch are indelible. Joel Green, David Ruskin, Christina Kim, Kristina Wilson, and Benedict Williams have shown me that, as a grad student, I am in exquisite company. Jess Druszkowski, Jed Lackritz, Amelia Harnas, Shann Chaudhry, and many others, have befriended and loved me despite the fact that I am a graduate student. Mentors at Pomona College and Pitzer College, including Jay Atlas, Robert Thornton, Mary Paster, Carmen Fought, and John Fought, have given me intellectual reference and haven, and have fed me more than once. Cody
Tucker, Tom Obert and Sabra Wood have been supportive forces outside of the academy. The skaters, volunteers, and fans of Roc City Roller Derby have challenged parts of my mind and body that I had forgotten about, and I will always be grateful that I had the chance to be an athlete and a scientist at the same time. Katie Carbary has been my life raft, both within and outside of the lab, and I refuse to imagine what grad school would have been like without her. My family, especially Terri Klein, Stephan Klein, and Ila Burgardt, have never stopped providing me with support and freedom to learn and take risks with a state-of-the-art safety net below.
Abstract

Language refers to ideas, objects, and events in the environment and mind. Because linguistic choices are affected by mental state and context, language provides clues to how we perceive and remember. Psycholinguists aim to understand links between referential world, linguistic form, and mental representation.

One ubiquitous construction is the definite noun phrase, which marks a unique referent in discourse. However, Carlson and Sussman (2005) observe special nouns, weak definites, where this generalization fails. “Bill heard about the riot on the radio, and Mary did too,” necessitates that Bill and Mary heard about the same riot, but not necessarily via the same radio. Roberts (2003) describes “informational definiteness,” which may be reconcilable with these NPs; she also asserts that psychological approaches can shed light on definiteness.

In this dissertation, I experimentally examine the relationship between weak definite NPs, referents, and conceptual knowledge. I aim to 1) assess claims about weak definites, 2) understand their comprehension and mental representation, and 3) broaden linguistic and psychological insight about informational definiteness.

Five experiments are presented that address central claims about weak definites and take unique approaches to understanding the language-cognitive link.

Experiment One investigates uniqueness presuppositions and weak definites. I elicit referential interpretations with a storyboard task and demonstrate that weak definites lack unique reference.

Experiment Two assesses whether weak definites establish novel discourse referents. The task is a picture/sentence judgment study, where participants rate the
semantic fit of text to illustration, and results suggest weak definites do not establish discourse entities.

Experiments Three-A and Three-B show that weak definites convey enrichment based on world knowledge of conventions and concepts by measuring conventionality judgments in a visualization task.

Experiment Four creates new weak-definite-capable nouns in the lexicon by manipulating semantic properties of novel words and referents. The results help explain the functional role weak definites play in the lexicon.

I conclude that weak definite NPs are not truly referential, but are linguistic “shortcuts” that evoke rich representations of conventional activity types, and that new weak definites can appear given relevant conceptual properties like interchangeability and similarity across contexts.
# Table of Contents

Curriculum Vitae ii

Acknowledgments iii

Abstract v

List of Tables x

List of Figures xi

Foreword 1

1 Introduction 2

1.1 Informational Definiteness .............................. 4

1.2 Processing Definiteness ................................. 7

1.3 Weak Definite Noun Phrases ............................. 11

1.4 Overview of Experiments ............................... 19

2 Weak Reference and Uniqueness 21

3 Weak Definites and the Introduction of Discourse Referents 33

3.1 Experiment Two ........................................ 34
# TABLE OF CONTENTS

4 Semantic Enrichment 41  
4.1 Experiment Three-A ................................. 41  
4.2 Experiment Three-B ................................. 48  

5 Categories and Semantic Features 51  
5.1 Experiment Four ................................. 51  

6 Conclusion 60  
6.1 Experimental Summary ............................. 60  
6.2 Conceptual Knowledge and Weak Nouns ............................. 62  
6.3 Unresolved Questions about Articles ............................. 64  
6.4 Referentiality ................................. 68  

Bibliography 70  

A Experiment 1 Materials 75  

B Experiment 2 Materials 77  

C Experiment Three Materials 79  

D Experiment 4 Materials 81  
D.1 Artificial Lexicon Science Fiction Story ....................... 81  
D.2 Comprehension Questions ............................. 98  
D.3 Experimental Questions ............................... 101  

E Incremental Interpretation of Chinese Classifiers and Massifiers 105  
E.1 Background ........................................ 105  
E.2 Introduction ........................................ 105
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Mean Ratings for One-Referent Scenes</td>
<td>37</td>
</tr>
<tr>
<td>3.2</td>
<td>Mean Ratings for Two-Referent Scenes</td>
<td>37</td>
</tr>
<tr>
<td>3.3</td>
<td>Three-way ANOVA for Experiment Two</td>
<td>38</td>
</tr>
<tr>
<td>3.4</td>
<td>Two-way ANOVA for Two-Referent Conditions</td>
<td>39</td>
</tr>
<tr>
<td>4.1</td>
<td>Experiment Three-A: Yes/No Responses</td>
<td>46</td>
</tr>
<tr>
<td>4.2</td>
<td>Experiment Three-B: Yes/No Responses</td>
<td>49</td>
</tr>
</tbody>
</table>
List of Figures

2.1 Experiment One Setup ........................................... 23
2.2 Experiment One in Action ........................................ 31
2.3 Experiment One Results ........................................... 32
3.1 Experiment Two Setup ............................................ 35
3.2 Experiment Two Results ........................................... 40
6.1 Weak Definites in Natural Contexts .............................. 63
E.1 English Measure Phrases Visual World Computer Display .... 109
E.2 General English Measure Phrase Eye-Movement Proportions .... 112
E.3 Specific English Measure Phrase Eye-Movement Proportions .... 113
E.4 English Measure Phrase Target Looks ............................. 114
E.5 Chinese Massifiers Visual World Computer Display ............ 117
E.6 General Chinese Massifiers Eye-Movement Proportions ........ 119
E.7 Specific Chinese Massifiers Eye-Movement Proportions ........ 120
E.8 Count Classifier Visual World Computer Display .............. 123
E.9 General Chinese Classifiers Eye-Movement Proportions ........ 124
E.10 Specific Chinese Classifiers Eye-Movement Proportions ....... 125
E.11 Comparing Massifier Referent Selection ......................... 128
E.12 Comparing Classifier Referent Selection ......................... 129
Foreword

I am the primary author of the entire text of this dissertation. Several collaborators helped with portions of this research and have coauthored related manuscripts for publication. These collaborators include Rachel Sussman, Greg Carlson, Mike Tanenhaus, and especially Whitney Gegg-Harrison.

Experiment One: Sussman helped create experimental items. Gegg-Harrison helped run subjects and implement the mixed-effects linear regression models.

Experiment Two: Carlson and Gegg-Harrison helped create items, and Gegg-Harrison programmed the study and helped analyze data.

Experiments Three-A and Three-B: Tanenhaus and Gegg-Harrison helped design experiment and analyze data.

Experiment Four: Undergraduate researchers in the Tanenhaus lab coded subject responses for post hoc analyses, which Gegg-Harrison implemented.

Materials presented in Appendix E represent a collaborative project with Renjie Li, Florian Jaeger, Greg Carlson, and Mike Tanenhaus.

Overlapping research has been published in Carlson et al. (2006), Klein et al. (2009), and Klein et al. (in press).
1 Introduction

Language, as communicative tool, frequently refers to ideas, objects, and events in the environment and in the mind. These entities can be named, discussed, evoked, alluded to, etc., in accordance with the intentions of the speaker and her knowledge and beliefs about her reality. Linguistic communication, then, is among the most rich and concrete set of clues to how we perceive and remember the world around us. While speakers have abundant choice in what they say at any given time, the correspondence between speech, thought, and the environment is not arbitrary. One aim of psycholinguistics is to understand this mapping between referential world, linguistic form, and mental representation, and how this correspondence affects information processing.

There remains a somewhat impoverished understanding of the mental underpinnings of linguistic reference, despite its ubiquitousness. What exactly are people thinking about, and how do they express that in words? The malleable nature of the referential domain (a sort of “ambient linguistic array”)—which changes along with a speaker’s visual, conversational, and mental environment—makes understanding the relationship between words and mental representations particularly challenging.

There are clear cases where the label a speaker uses to refer to someone or something intuitively reflects the speaker’s state of mind. People, for instance, can be referred to with a variety of labels, including names, titles, and pronouns, and choice of a label can convey formality, relationship, or intimacy. A roller derby
skater may be “Muffy Stopheles,” “the green jammer,” “our captain,” “her,” etc., and each label reveals something about the speaker’s relationship to this skater. Linguistic form can also vary depending on what else is present in the environment. For example, “the whistle” might become “the pink whistle” or “Jack’s whistle” if more than one whistle is in the environment. These examples with nominals clearly illustrate that the speaker’s perspective and world knowledge, at a minimum, have some relationship to the form reference takes.

A more subtle way many languages have of marking distinctions in referential prominence is through the use of articles or determiners, and I will focus on this particular feature in this paper. Canonically, the definite article is analyzed as presupposing unique existence of an entity, with respect to the common ground of the interlocutors (and not, for example, the universe) (Stalnaker, 1974). In other words, a definite noun phrase refers to something that is assumed to exist uniquely in the relevant domain of discourse. In contrast, this is typically not true of indefinite noun phrases, which indicate a novel or non-unique entity, or one that is known by the speaker to exist, but is not assumed to be in the common ground shared by both speaker and addressee (Heim, 1982). However, despite the ostensible systematicity of article use, the true mental impact of these and other function words has remained relatively elusive to psychologists. Thus, a word such as “the,” which is the most frequent word in English\(^1\), is a word we all use and understand, but its relationship to mental representation is not entirely clear.

Further complicating the issue is the existence of definite noun phrases that seem inconsistent with even the most intuitive generalizations about the usage of the definite article. The particular cases I examine will be weak definite noun phrases, which, unlike typical definites, do not uniquely refer. Example (1) demonstrates such a case where the definite NP, “the hospital,” serves as an antecedent for both “Highland” and “Strong.” This can be contrasted with a syntactically identical paragraph, as in Example (2), where the second sentence seems to contradict the first.

---

\(^1\)Data cited herein have been extracted from the British National Corpus, distributed by Oxford University Computing Services on behalf of the BNC Consortium. All rights in the texts cited are reserved.
(1) After a particularly serious accident, two skaters went to the hospital. Mary went to Highland and Josie went to Strong.

(2) *After a particularly serious accident, two skaters went to the chapel. Mary went to St. Claire and Josie went to White Lotus.

In this thesis, I will examine the properties of weakly referential definite noun phrases, or weak definites. In a series of four experiments, I will test theoretical claims about weak definites’ lack of uniqueness presuppositions and their dissimilarity both to traditional definites and *indefinites* as well. Then I will present the first experimental data that reveal weak definites to be semantic shortcuts for much richer mental representations. Finally, I will demonstrate experimentally that weak definites are an open class, and that the ability of a noun to allow a weakly referential interpretation is directly linked to conceptual and semantic features of the noun. In tandem, these results reveal new information about the psychological correlates of linguistic definiteness, introduce novel experimental paradigms for psycholinguistics, and identify prominent conceptual features that influence linguistic form and interpretation.

### 1.1 Informational Definiteness

Linguists have debated the nature of definiteness in English. Some scholars have argued that definites presuppose semantic uniqueness (Russell, 1905, inter alios), while others claim definites presuppose familiarity (Heim, 1982, for example), and still others posit that definites neither require nor are satisfied by either uniqueness or familiarity (Birner and Ward, 1994, for example). The link between semantic uniqueness and definiteness appears quite strong in examples like (3), where use of an indefinite article is infelicitous when the referent is known or assumed to be unique in the world. A phrase such as “the love of her life” can sensibly only refer to one individual, and so it is definite by virtue of its singularity in the world. But the definite article is clearly found in many situations where the noun in question
is only unique with respect to some particular domain, and not in the universe, as illustrated in Example (4). The reader knows that no “team” is not unique in the universe. In this case, definiteness must operate with respect to some known domain. In other words, definiteness corresponds more broadly to uniqueness in the discourse or common ground than it does to true uniqueness in the world.

(3)  *Liza just met a love of her life.

(4)  Stella recently joined the team.

Familiarity-based analyses of definiteness capture intuitions that previously mentioned entities, having become part of the common ground of discourse, where common ground is roughly ”the background beliefs or assumptions which are shared by the speaker and his audience, and which are recognized by them to be so shared” (Stalnaker, 1974, p. 48). Entities that are part of the common ground can felicitously appear in definite form, as in Example (5), where “helmet” is first introduced with an indefinite article, but then is referred to on repetition with a definite NP. Many other examples can be found, however, where the definite article appears with nouns that are neither semantically unique nor have appeared in prior discourse, as in Example (6).

(5)  Thea ordered a helmet and mouthguard online. The helmet turned out to be too small.

(6)  Thea ordered all her roller derby gear online. The helmet turned out to be too small.

Craige Roberts (2003) addresses cases like this by making a distinction between strong familiarity, in which entities have been overtly introduced into the common ground, and weak familiarity, in which entities are simply entailed by the common ground. This allows for the use of definites as in Example (6), where “helmet” has not appeared overtly in the discourse previous to use in a definite NP. This is acceptable, argues Roberts, as long as the existence of the helmet is entailed by
mutual knowledge; in this case, “roller derby gear” entails “helmet,” which licenses it to appear definite in the second sentence. Unmentioned but entailed referents are described by Roberts as being weakly familiar.

Roberts posits the idea of informational definiteness to describe uniqueness “relative to a description of a discourse referent, not an entity in the model” (Roberts, 2003, p. 328). It is this type of uniqueness – which requires at least weak familiarity but does not necessitate specificity – that Roberts attributes to definite NPs in English. To summarize, in Roberts’s theory of informational definiteness, no earmarked or specific entity need be represented in the minds of interlocutors for definiteness to be licensed: a unique description of a referent relative to the discourse context and entailed by the common ground opens the door for definiteness. This can be captured in (7), which I have adapted from her argument.

(7) Marnie decided to put outdoor wheels on her skates. She bought nine wheels in case one went missing or was defective, but she hopes she will never need the extra wheel.

In Example (7), none of the nine wheels in particular is earmarked as being “the extra wheel,” since they are all indistinguishable. But under Roberts’s theory, the common ground knowledge that quad skates have eight wheels and Marnie has nine wheels, which means exactly one wheel can be classified as “extra,” is enough for the definite phrase to felicitously appear. All of the nine wheels are interchangeable, so no specific wheel sticks out as being the extra one; however, comprehenders can deduce that one wheel will end up becoming the spare, since the other eight will be bolted to Marnie’s skates. I will return to the role of interchangeability in definiteness later in this thesis, but for now it is important to note that in Roberts’s framework, common ground knowledge about the identifiability of some particular item in the referential domain allows for definite NPs to appear.

Outlining the distinction between weak and strong familiarity raises questions about the mental processes involved in assessing common ground entailments. Familiarity can license definiteness, and it cannot be assessed independently from the
mind of the speaker and comprehender. Entailment relations, in everyday language use, also appear dependent on interlocutors’ world knowledge, as well as memory of prior discourse (Horton and Gerrig, 2005). Thus how interlocutors mentally establish common ground—which may contain information about shared past experiences, current physical environment, or presumed community co-membership—is a processing-based question that moves from theoretical accounts of definiteness and toward mechanistic and psychological approaches to understanding this problem.

1.2 Processing Definiteness

Assessing the common ground and linking that knowledge back to language and linguistic form necessitates a rapid mental ability to integrate linguistic input and contextual information. Fortunately, there is evidence that humans do capably engage in this type of information processing. Psycholinguistic research has demonstrated that as linguistic information arrives, listeners both rapidly integrate new input with preceding context and generate expectations about what may come next. Some of the most striking evidence of this incrementality comes from studies using the visual world paradigm (Cooper, 1974; Tanenhaus et al., 1995). In this approach, researchers track a language user’s eye-movements as she hears an utterance related to a visual scene. This allows the experimenter to examine where a listener is looking, millisecond by millisecond, as a speech stream unfolds. The position of eye-gaze is influenced by the attention and thoughts of the comprehender, and so this technique can reveal the way in which subjects resolve temporary ambiguities in language or anticipate certain scenarios, given different contextual conditions.

Visual world research has demonstrated the effect of many different factors on language comprehension. The contextual information that listeners can use includes low-level information in the speech stream, such as phonemic coarticulation, syntactic cues, and lexical semantic information that draws on world knowledge. For example, in a seminal study, Altmann and Kamide (1999) found that as listeners hear a verb, they generate anticipatory eye-movements to a likely theme or patient
(Boland and Blodgett, 2001, discuss related factors). In the presence of a verb such as “eat,” comprehenders immediately begin to preferentially fixate objects in the visual display that are edible, such as “cake.”

Similarly, Chambers, Tanenhaus, Eberhard, Filip, and Carlson (2002) have demonstrated anticipatory eye-movements to objects that contain the appropriate affordances for certain function words like prepositions: subjects look at a tin can rather than a plate when the preposition “inside” is used. This strongly suggests that knowledge about items in the environment and how they function, as well as the linguistic conventions that describe such usage, is readily available to comprehenders.

One puzzle given these sorts of effects is the nature of the comprehension of definite articles, where current online processing data is equivocal and does not seem to capture the strength of offline judgments and theoretical accounts of definiteness. Intuitively, definiteness matters to the way language, specifically reference, is understood and should therefore impact reference resolution and sentence processing. And in fact, in visual world studies there is clear evidence that listeners make some use of information provided by a definite article. For example, in a display with two jars, listeners show signs of confusion when there are two potential destinations for an instruction that uses a definite noun phrase (e.g. “put the cube in the jar” when two jars are present) but not when the instruction uses an indefinite noun phrase (Chambers et al., 2002). Hearing a definite when no compatible referent had any reason to be seen as unique slowed the comprehension process, which indicates that listeners assess online the compatibility of linguistic form relative to the scenario being described. Listeners may be making predictions about upcoming words based on uniqueness conditions, they may be matching the utterance to the scene, or they may be engaging in a complex synthesis where different weight is given to multiple information sources; in any case, the mismatch between hearing a definite NP while not knowing of a compatible referent that is unique or familiar becomes a problem for the listener. Hearing a definite article can have a rapid (negative) impact on sentence processing, in other words, when that definite NP is inappropriate
in the referential domain.

However, despite this, it does not appear that listeners make anticipatory eye-movements to a uniquely identifiable referent as soon as they hear a definite article. This lack of positive evidence is unexpected, given the body of research showing that comprehenders rapidly generate expectations based on other linguistic sources during sentence processing. In the Chambers et al. (2002) studies there was often one unique potential container (e.g. a bowl) and two other potential containers of the same type (e.g. both a large and small can). Yet, listeners showed no tendency to preferentially fixate the unique container (the bowl) upon hearing “put the cube in the...”. All containers remained candidates, even though the bowl was the only unique container in the scene. Whereas unlicensed definiteness can be disruptive to language comprehension, it is less clear that linguistic cues about definiteness can be immediately used to generate expectations about what the listener will hear next.

Why might listeners neglect the definite article in circumstances where it could be informative? One possibility is that articles are short and tend not to receive prosodic stress. Shorter, less prominent words in the speech stream may fail to engage the listener’s attention. However, there is considerable evidence that listeners make immediate use of much more subtle and fine-grained differences in the speech signal. For example, visual world studies have found that listeners make immediate use of within-category differences in voice onset time (McMurray et al., 2002) and can even use such information prior to subsequent information about vowel duration (McMurray et al., 2008). In addition, listeners make immediate use of coarticulatory information in vowels (Dahan et al., 2001; Dahan and Tanenhaus, 2004), and they use vowel duration to distinguish between input that is phonemically consistent with a monosyllabic “embedded” word such as “ham” and a polysyllabic carrier word such as “hamster” (Salverda et al., 2003, 2007). Thus it seems unlikely that listeners would not have sufficient phonetic information available as they heard a definite article.

It also seems unlikely that listeners would ignore uniqueness information associated with the article out of a need to wait for a noun to determine the type
of unique entity under discussion. Listeners are clearly able to use the implicit contrast associated with pre-nominal scalar adjectives as they hear the adjective (Sedivy et al., 1999; Heller et al., 2008, for example). For example, upon hearing “the tall...,” comprehenders anticipate not the tallest potential referent but an object that has a shorter competitor also present in the scene, for example, the taller of two candles rather than a (taller) pitcher. In the case of scalar adjectives, listeners are expecting a referent that is the member of a contrast set, and they demonstrate this expectation before they process any phonetic information from the noun. This use of contextual and pragmatic information during reference resolution makes it seem surprising that uniqueness presuppositions from the definite article could not impact sentence processing at the same rapid rate.

It is possible, however, that the presence of a definite article does not, on its own, actually provide strong uniqueness constraints. Adjectives and adverbs can directly follow the definite article in English, and those modifiers may change the relevant dimension of uniqueness with which the article corresponds. That is, words subsequent to the article, such as the adjective “tall” in “the tall candle,” could actually suggest that there are multiple candles (one taller than the other). In this case, the definite article highlights a size property unique to one particular candle compared to other candles, and does not require that there only be one candle. In Examples such as this, the listener could be “garden-pathed” if she anticipated a singleton. Considering the frequency of such grammatical constructions, it may be more prudent for a listener to wait for more information rather than generate expectations from an article alone.

Since language users are adept at integrating phonemic, sub-phonemic, semantic, and contextual information during the earliest moments of comprehension, the failure to use a definiteness cue online suggests that definiteness may be associated with complex and context-dependent mental representations and calculations. As I have discussed, Roberts (2003) posits an analysis of definiteness that allows for weak familiarity and acknowledges that informational definiteness is determined by mental processes. Examinations of canonical definite NPs within the visual world
paradigm have not yet painted a clear picture of these psychological correlates of definiteness. The importance Roberts gives to weak familiarity critically suggests that conceptual knowledge and experience can impact judgments about definiteness. When a listener and speaker both know that every car has exactly one steering wheel, for example, then introducing a car into the discourse consequentially allows a definite such as “the steering wheel” to be used and accepted. Thus a better understanding of how the definite article is comprehended requires some probing into the relationship between language and conceptual knowledge.

I am attempting to take a first step in this direction by examining a class of definite noun phrases that ostensibly do not conform to canonical uniqueness or familiarity accounts. By looking at the most unlikely occurrences of definite NPs—ones in which the noun does not seem unique along any dimension—I hope to find patterns that reveal something about definiteness more generally. In the process, I further hope to offer some reconciliation between these puzzling constructions and analyses of linguistic definiteness. Before proceeding with my experimental approaches to this problem, I wish to introduce the reader to the phenomenon of interest, weak definite noun phrases.

### 1.3 Weak Definite Noun Phrases

Many occurrences of common noun phrases beginning with the definite article “the,” such as “the hospital,” do not appear to necessitate associated uniqueness presuppositions (Barker, 2005; Birner and Ward, 1994; Carlson and Sussman, 2005). In certain contexts these phrases can appear more semantically indefinite in character than definite. Following Poesio (1994) and others, I will refer to these anomalous definite noun phrases as weak definites.

Consider the following example from Carlson and Sussman (2005):

(8) Mary heard about the riot on the radio, and Bill did too.
Intuition suggests that Mary and Bill must have heard about the same riot, but not necessarily by means of the same radio. In a world where Mary and Bill were listening to two different radios in two different places, or even to two completely different broadcasts, this sentence still seems acceptable. The phrase, the radio, is one example of a weak definite NP, because mysteriously, the definite article preceding “radio” does not seem to require that this entity be unique, although it certainly allows for that possibility. While it is certainly possible, given this sentence, that Mary and Bill heard about the riot on the same radio, it is not necessary; weak definites thus appear systematically ambiguous between a “regular” interpretation, in which there is (a presupposition of) a uniquely identifiable referent, and a “weak” interpretation, in which there is not.

To a first approximation, weak definite readings seem to be restricted to particular nouns in the lexicon. For example, although radio allows a weak interpretation in a definite NP in Example (8), this is not necessarily true of synonyms or close lexical associates. The following pair of VP-ellipsis sentences illustrate the lexically restrictive nature of weak definites:

(9) Benedict listened to the radio, and Lance did too.
(10) Benedict listened to the stereo, and Lance did too.

Example (9) seems to allow a weak interpretation, while Example (10) does not. I have designed Experiment One to test this claim empirically that weak definites do not carry a uniqueness presupposition in online sentence comprehension.

One plausible hypothesis is that these weak definites are, at their core, indefinite noun phrases that happen to have a definite article, perhaps through some artifact or accident of our language’s history. The following dialogue serves as an illustration of common intuitions about unique reference in definites (“the arena”), indefinites (“an arena”), and weak definites (“the hospital”):

(11) Sarah: Where did they take the hurricane victims?
     Otto: To the arena/ an arena/ the hospital.
Sarah: *Which one?*
Otto: *I don’t know.*

Sarah asks Otto where victims of a hurricane have been transported. In response to Sarah’s question, Otto gives a place-noun, which is either a canonical definite, an indefinite, or a weak definite NP. When asked for more specificity, Otto replies with uncertainty. Otto’s uncertain answer, *I don’t know*, is a perfectly acceptable response to Sarah’s question, *Which one?*, if Otto originally named an indefinite place, *an arena*. Otto knows the victims went to some arena, but he does not know which one, and his answer reflects this mental state without raising linguistic hackles.

If Otto named a place with a canonical definite NP, such as *the arena*, it no longer seems acceptable when Sarah asks *Which one?* for him to reply that he does not know. Otto’s use of the definite article has signaled that he is referring to a known, particular arena. Contradicting that presupposition later comes across as infelicitous.

Interestingly, with a weak definite such as *the hospital*, Otto’s eventual uncertainty about which particular hospital does not seem problematic. Despite the appearance of definiteness, Otto felicitously utters that the victims were taken to *the hospital* even when he does not know which one; this uncertainty is infelicitous with a regular definite like *the arena* and suggests that weak definites may simply be indefinites with an irregular form. Experiment Two investigates whether there are differences between weak definites and indefinites in referential contexts.

Although weak definites appear to be truth-conditionally similar to indefinites in examples like (11), there is still the mystery of why some nouns would allow for weak uses of the definite article while others do not. There is also the puzzle of why these nouns, which could appear with an indefinite, often do not. Furthermore, if experimental results demonstrate that weak definites behave neither like conventional definites nor like indefinites, questions arise about what communicative purpose they may have. In fact, on closer scrutiny, weak definites seem to convey information beyond that of a typical definite or indefinite noun phrase. Al-
though the influence of function words on the earliest moments of language comprehension is equivocal, and weak definite interpretations are perhaps subtle and context-sensitive, there may be cause to consider them semantically empowered, as we see when comparing the preferred readings for a weak definite like (12) and an indefinite like (13):

(12) Izzy is in the hospital. → Izzy is assumed to be a patient getting medical treatment.

(13) Izzy is in a hospital. → Izzy is inside a hospital building in some unknown capacity.

The example in (12) has a dual reading, which is the case with weak definites, since they are identical in form to canonical definites. The first interpretation of (12) is that of a typical definite NP, where “the hospital” refers to a particular hospital in the common ground, and Izzy is at that known location. The other reading involves “in the hospital” being interpreted weakly, without reference to a specific hospital. Critically, while this reading of Example (12) does not presume a unique hospital, it does seem to convey something richer about why Izzy is in a hospital—to receive medical treatment. The weak definite appears to be a natural shorthand for a richer scenario.

Notably, the indefinite in (13) does not appear to convey the same level of enriched meaning as the weak definite. This suggests a possible hypothesis: rather than referring to an entity as regular definites do, or establishing a discourse entity as indefinites do, weak definites potentially evoke a conventional activity or event type for the hearer. Common conceptual knowledge could in fact be made salient by the presence of a weak definite NP. Experiments Three-A and Three-B examine this hypothesis and probes the strength of these enriched interpretations and the extent to which they are part of the meaning of weak definites. Experiment Four explicitly manipulates conceptual properties and measures the effect of such manipulations on the acceptability of using a definite article “weakly.”
1.3.1 Contextual Environments of Weak Definites

The label “weak definite noun phrase” is somewhat misleading in that it does not capture the critical role linguistic context plays in making the weak reading available. It is tempting to attribute weak referentiality to the noun alone, since the noun sticks out as a linguistic anomaly. However, contextual information becomes critical in licensing weakly referential phrases. Verb use and modification in particular can determine the availability of a weak interpretation. Modification of the noun seems to emphasize the referentiality and unique identifiability of the noun and must typically be absent in order for a weak interpretation to be available. Otherwise, uniqueness presuppositions prevail, as in the following examples with “bank” and “phone,” both of which can be found in weak definite NPs. In these examples, the elided phrase in the second conjunct is coreferential with the definite NP in the first conjunct, and no weak interpretation is available, due to the presence of modifiers “new” and “red”:

(14) *Barry went to the new bank today, and Farrah did too.*

(15) *Emmeline talked on the red phone, and so did Eva.*

Notice, however, that expressive modification, which does not highlight properties of the referent but provides editorial commentary on the nature of the event, still allows weak interpretation, as in Example (16). This suggests that it is something conceptual, and not syntactic, about modification that can require the uniqueness of a canonical definite:

(16) *Candi is in the friggin’ hospital, and so is Jessi.*

The contextual information of the action or event being described also seems critical to licensing a weak reading. Preceding information, such as verbs, must not be inconsistent with the conventional activity where a particular weak definite is appropriate. If a verb preceding a weak noun describes an event contradictory to the
typical enrichment weak interpretations convey, then the NP seems to lose its weak interpretation; only a regular, unique interpretation is available in these cases. This is demonstrated more clearly by the contrast between (17), which allows a weak interpretation, and (18), which necessitates that a specific bus has been hijacked. If the intended sense of “took” was “stole” and not “rode,” then “took” would pair with “hijacked” in requiring reference to a uniquely identifiable bus. As long as the intended meaning is “rode,” which is the prototypical activity associated with buses, “the bus” need not refer to any specific vehicle.

(17) \textit{Linus took the bus.}

(18) \textit{Linus hijacked the bus.}

Changing the verb from “took” to “hijacked” forces “the bus” to refer to a particular entity. When Linus “took the bus,” the specific vehicle involved does not seem important. He may have even transferred and ridden multiple buses to get to his final destination, and this would be compatible with the sentence in Example (17). However, when a less canonical bus-related activity such as a hijacking is described, a weak interpretation of the NP with an enriched meaning is not available.

That this governing restriction on weak reference hinges on semantic content highlights the importance of the conventionality of weak definite NPs. The relationship between the verb, the NP, and the actual activity being described is non-trivial to allowing weak interpretations. World knowledge, once again, appears to play a strong role in the licensing of weak definites, and in their ultimate interpretation to the listener. The event described must be common and conventional in order for the NP to lack uniqueness presuppositions.

This fragility may lead to the objection that weak definites are simply a type of idiom, where a closed set of words and phrases lead to a richer meaning than their literal components would suggest. Like idioms, the particular identity of the noun is critical to accessing the enriched meaning. However, in the case of weak definites, synonymous verbs often can be interchanged while preserving the enriched and
non-unique interpretation, as in Example (19). This is not the case with idioms, as in Example (20): although “nibbled” is conceptually similar to “bit,” “nibbled the bullet” means something very different than “bit the bullet.”

(19)  
*Linus took/rode/caught the bus.*

(20)  
*Roxy nibbled/chewed/munched the bullet.*

Weak reference is allowed when the verb sense, regardless of the particular word used, is compatible with a canonical activity type. This means that weak definite phrases, while possibly greater than their literal sum, have a mapping to the world that I argue is systematic and calculable from conceptual knowledge. With idioms, however, a particular verb must be used in a set phrase, and the meaning is not at all transparent or calculable from the individual words involved. That is, “bit the bullet” need not involve teeth or ammunition, but “took the bus” must involve at least one bus, even though the bus need not be specified or familiar.

It should be emphasized again that, even when a weak interpretation is licensed, a regular reading of the definite NP would look identical, rendering many of these NPs ambiguous or underspecified with respect to weak referentiality in the absence of context. Examples like (19), which allow a weak reading, could also refer uniquely to one particular “bus” within the common ground but they need not. This overlap of form makes it difficult to isolate and investigate weak definite NPs, and it also leads to one of my key hypotheses: weak definite noun phrases are critically reliant on more than the noun. The difficulty in identifying or creating definite NPs that unambiguously have a weak interpretation becomes important in an experimental setting. This ambiguity is a feature of weak definites that I will address in later chapters about the possible origins and purpose of such phrases.

### 1.3.2 Across Languages and Within Categories

While I limit the current set of experiments to English weak definites, these structures are by no means limited to English or Germanic languages. Many languages
with articles have special cases where a definite NP does not uniquely refer. Like English, Dutch and German have such cases, but so do Romance, Basque, and even Celtic languages.

Across languages and dialects, weak definite usage often appears in complementary distribution with bare nominals, another structure where articles behave oddly and which can be interpreted in much the same way as weak definites. For example, (21) is an American English bare nominal expression that does not require unique reference, whereas the Hiberno-English dialect equivalent in Example (22) is a weak definite. In Irish, the equivalent phrase in Example (23), “ar an teilifís,” is also a weak definite. In all three cases, Jess is featured in some kind of televised programming, and the particular tube displaying her broadcast image is not something mutually known, or even relevant, to speaker and addressee.

(21)  

Jess is on TV.

(22)  

Jess is on the box/telly.

(23)  

Tá Jess ar an teilifís.

Similar cases appear in German with contracted preposition/determiner phrases, such as in Examples (24) and (25), “Maria went to the supermarket, and Hans did too,” with a non-contracted and contracted form of the prepositional phrase, respectively. The non-contracted forms have the ability to function as a topic of the sentence. In other words, Example (24) can refer to a specific market deictically. By contrast, the contracted form ostensibly lacks anaphoric force and is backgrounded in the discourse (Waldmüller, 2008). Interesting for our purposes is the observation in Schwartz (2009) that weak interpretations can only occur with the contracted form, as in Example (25), and not with the full form in Example (24). Although it is not within the scope of this dissertation, the link between discourse backgrounding and weak interpretations that is demonstrated in the case of German preposition-determiner contractions may be a venue for future research.
My motivation for seeking examples of weak definites (and interpretationally similar bare nouns) in other dialects and languages is that cross-linguistic cases like these might indicate conceptual universals. One might ask, for example, whether there are non-linguistic properties of these phrases that are consistent and independent of a particular lexicon. If weak definites are merely lexical anomalies, they could have arisen with nearly any type of noun: I would not expect them to arise systematically in the same contexts or have the same types of enriched meanings within a language, and certainly not between very distinct languages.

In fact, when weak definites appear across languages and dialects, they do often turn up in the same conceptual categories. These include mass transit, mass communications, institutions, habits and chores, interchangeable parts and places, and special skills. That weak definites are found in similar conceptual space across a variety of languages and language families suggests they are not arbitrary exceptions, but instead reflect something meaningful about our world knowledge and experience. This cross-linguistic commonality hints that weak definites may interact with conceptual knowledge in a way that may, if appropriately examined, shed light on our knowledge of definiteness and the mental representation of language. Experiment Four explicitly addresses the relationship between concepts and weak reference in more depth by attempting to create new weak definites in an artificial language by manipulating conceptual and semantic information.

1.4 Overview of Experiments

Weak definite noun phrases can be examined systematically using behavioral psycholinguistic tools. My primary aim is to apply such tools in order to 1) assess theoretical claims about weak definite noun phrases in particular, 2) understand the way such phrases are comprehended and mentally represented, and 3) broaden our
insights about what informational definiteness entails and how it becomes expressed linguistically.

Five experiments will be presented in the chapters to follow. Each addresses one theoretical claim about weak definite noun phrases and takes a unique approach to understanding the link between such phrases and more general cognitive mechanisms.

Experiment One investigates the claim that, unlike typical definites, weak definites lack a uniqueness presupposition. I elicit referential interpretations with a storyboard action task to test this claim experimentally.

Experiment Two assesses whether weak definites are interpreted as establishing novel discourse referents the way canonical indefinites can. The task is a picture/sentence judgment study, where participants rate the semantic fit of text to illustration.

Experiments Three-A and Three-B empirically tests the claim that weak definite NPs convey a level of enriched meaning that draws on world-based knowledge of conventions and concepts by measuring conventionality judgments in a visualization task.

Experiment Four attempts to systematically create new weak-definite-capable nouns in the lexicon by manipulating semantic properties of novel words and referents. This would help explain the functional role weak definite NPs play in the lexicon.
2 Weak Reference and Uniqueness

Perhaps the most central claim about weak definites is their ostensible lack of unique reference to an entity in the common ground. Intuitions that weak definite noun phrases do not presuppose uniqueness are what set these phrases apart from ordinary definites. Experiment One was designed to explicitly evaluate the claim that weak definites are not uniquely referential in comprehension.

One approach to exploring this problem is to examine how language users map weak definites onto entities in the world when they are asked to choose which item is the best referent for a specific utterance presented by the experimenter. Assessing comprehenders’ referential selection under experimental conditions could reveal critical information about how weak definites convey information.

The challenge of designing such an experiment is that weak definite NPs are identical in form to their regular definite counterparts. In sentences involving VP-ellipsis (such as Example (1)), the stronger co-referential interpretation (where Muffy and Gretchen went to the same beach) is always felicitous, because these definite NPs are systematically ambiguous between a weak and a regular interpretation. While Gretchen and Muffy need not have gone to the same beach, in the absence of more information (e.g. knowledge that they are on opposite coasts), the unique definite reading is not excluded by the form of the NP. This ambiguity of form that is inherently part of my phenomenon of interest must be considered in the design of any tasks or stimuli meant to reveal the way such phrases are understood by language users.
Another important thing to note is that actions in a behavioral task become easier when repeated. Since weak definites look just like their regular definite counterparts, and since I am examining anaphoric reference, subjects are potentially acting out repetitions of reference to the same objects. Moving one’s hand to the same point in space to grab the same object, in other words, becomes mentally easier over time, which may give co-referential selection (corresponding to unique interpretations of the definite NP) an advantage in laboratory tasks. Carlson et al. (2006) discusses an experimental task with weak definites where participants were able to click repeatedly on an identical referential token. The ease of this repetition potentially masked the salience of weak interpretations, which would license subjects to choose a new token rather than performing a repetition. Since the response corresponding to unique reference was easier than a response that would indicate weak reference, it is likely that this task underestimates how widespread and salient weak definites actually are. In other words, when subjects have just clicked on a “the newspaper” and have the option of clicking the same object in the same location, they have no reason to do otherwise, even if a weaker interpretation is available, since that weaker interpretation would require identifying a new token of “newspaper” and executing a new path of movement to select that token.

My study takes the cognitive ease and felicity of the stronger definite interpretation into account in its design, and aims to minimize the non-linguistic bias against possible weak interpretations by providing a laboratory environment where behavioral responses indicative of weak interpretations are no more difficult than those corresponding to strong interpretations. In other words, I attempt to level the behavioral playing-field between weak and canonical definite reference so that the former are not masked by task difficulty.

Following previous work by Brown-Schmidt and Tanenhaus (2008), which established that visual context encourages participants to treat differently demarcated regions as independent and cohesive discourse domains, I painted each half of a magnetic board a distinct color to evoke two visual domains. Each domain con-
CHAPTER 2. WEAK REFERENCE AND UNIQUENESS

Figure 2.1  Experiment One Setup: A two-toned magnetic board with a male and a female agent, and non-identical tokens of both the critical noun and the distractor noun.

tained magnets with images of a person, a non-identical token of a distracter object, and a non-identical token of the critical noun.

Figure 2.1 shows a cartoon version of the setup. Pre-recorded scenarios were constructed so that each scenario could be used with a weak definite noun (e.g. newspaper) and a comparable noun that does not have a weak reading (e.g. book), as in Example (1):

(2)  Rudy is a very literary guy. Today he wrote in his diary.
Then Rudy read the newspaper/book.
This afternoon, Patty read the newspaper/book too.
Participants were instructed to use the person and object magnets to act out the scenarios on the metal board. In the case of regular definites (e.g. book), I expected participants to interpret the first and second mentions of the critical NP as referring to the same object. If weak definites, as claimed, lack the uniqueness requirements of regular definites, then participants should be more likely to interpret the two NPs as referring to distinct tokens of the same object type. That is, subjects should choose the new “newspaper” the second time it is mentioned if “the newspaper” is a weak definite noun phrase.

2.0.1 Experiment One

Participants

Eighteen members of the University of Rochester community participated in the experiment for pay. Participants all had normal or corrected-to-normal vision and were native speakers of American English.

Design and Materials

The experiment had two conditions corresponding to the type of noun presented in the scenario: regular definite and weak definite. Twelve nouns with weak definite readings were chosen from examples in the literature and the intuitions of the experimenters. Because the task required objects to be easily identifiable, imageability of the noun was prioritized during stimulus creation. For each weak definite noun, a similar noun without a weak definite reading was chosen as its regular definite match. For example, the weak definite noun newspaper was paired with the regular definite noun book. A full list of stimuli can be found in the appendix.

A male speaker naive about the experimental purpose instructed to maintain a naturalistic and non-contrastive prosody was recorded reading short scenarios. For critical items, twelve scenario frames were created so that the critical noun could either be one of the twelve weak definite nouns, for experimental items, or its
regular definite counterpart, for control items. The first instruction in each scenario always referred to a distracter object in the visual display; the second instruction referred to the critical noun in a definite NP; the third repeats that critical NP as in Example (3):

(3)  Amos goes out a lot. This morning, he went to a cafe.  
     Tonight, Amos went to the movies/the concert.  
     Trish went to the movies/the concert too.

Half the items were imperative instructions to the subject, as in “Have her wear a hat.” The other half were short narratives that did not explicitly instruct the participant to carry out an action, and participants were told at the beginning of the experiment to act out all types of narratives. The narratives were recorded as three separate sound files, so that the experimenter could wait until the participant completed the first action before playing the next part of the story.

Two randomized lists and their backward counterparts were created, for a total of four lists. Each list contained four fillers and 12 scenarios, half involving weak definites and half involving regular definite control trials. The scenarios described magnetic clipart objects on a 40 cm by 50 cm metal board on an easel in front of the participant. Four metal boards were rotated throughout the study so that an experimenter could prepare trials in advance. Each half of metal boards were painted a different bright color (red, yellow, green, or blue), so that there were two visually salient color domains on each board.

Because of the visual salience of the color domains, and because each domain contained a token of both the distracter and the critical noun, practice trials and fillers were designed to counter the strong bias we found in piloting to treat each domain as a separate “universe” and avoid completely co-reference which might require switching domains. In other words, it was critical that participants knew they could move items back and forth between domains, so fillers that required such movement were included in the design. For example, a filler might refer to each character playing the game, where only one game token is present in the scene,
which would require movement of either the game magnet or a character magnet into the adjacent color domain. This implicitly indicated to participants that the characters and items on the metal boards were not bound to one particular color domain.

One filler used a weak definite to refer to an item that only had one token present in the scene. Another filler by contrast referred to a singleton with a regular definite. This ensured that the presence of only one token of an object in the display did not systematically correlate with our experimental manipulation. The remaining two fillers referred unambiguously to items present in the visual scene without using a definite NP.

**Procedure**

Participants were instructed to use the magnetic pictures to act out the narratives they heard. During all critical scenarios participants saw one male and one female character, two non-identical critical nouns of the same type (e.g. two different looking dogs), and two non-identical distracter nouns of the same type. One character, and one of each noun type were present on each half of the metal board, which we will refer to as a color domain. The precise position of the magnets was not systematically dictated, but the responsibility for placing the magnets on the metal boards in preparation for each trial belonged to a research assistant naive to the experimental manipulation.

Participants had three practice trials at the beginning of the experiment, and if they had no questions, the experiment began immediately following practice. Sessions were video recorded, so that referential choices could be analyzed later.

I predicted that, when hearing weak definite nouns like *newspaper*, participants would be more likely to pair each agent with his or her own token (e.g., each agent would have their own newspaper), whereas with regular definites, like *book*, participants would be more likely to move the old token from the agent in one color domain to the agent in the other. The dependent measure of Experiment One is the
proportion of times a participant selected the new token for the second mention of the critical noun during enactment of the final part of the narrative.

Despite our attempts to create an environment with fewer non-linguistic biases against the weak definite interpretation, I still had concerns that our task would under-represent the salience of weak interpretations, since in all cases, the strong interpretation was felicitous. The fact that weak definite NPs look exactly the same as canonical definites poses an experimental challenge: How will subjects know that the intended interpretation is weak and not strongly referential? Moreover, when participants have recently executed the motor planning involved in selecting the first token of the critical noun, selecting that token again would be very easy. Their hands and eyes had just located an initial referent. If strong definite interpretations are always available linguistically, and the visuo-haptic system finds the strong interpretation to be practiced and easy, this poses challenges for capturing weak interpretations in a lab setting. These lingering concerns led me to eye-track all participants during the task. I reasoned that, while the actual behavioral enactment might under-represent the weak interpretation, eye-movements could potentially provide evidence that the new token was being considered implicitly as a possible referent for weak but not regular NPs. Even if subjects always chose repetitions of the old token object, for instance, they may look at the new token as a potential candidate more often in the weak definite condition. If this was the case, I wanted to be able to capture this difference in the lab, which is why I chose to track participants’ eye-movements as they performed the task. It was because of this decision to eyetrack participants that I chose to repeat the full critical NP in my linguistic stimuli rather than use a VP-elided form, giving me lexical material to which I could time-lock the eye-movements if necessary.

2.0.2 Results and Discussion

I reviewed videotapes of each participant’s experimental session and coded each trial for whether the new or old token of the noun was selected upon its second mention in the critical utterance. By examining whether subjects interpreted repe-
tition of a (weak or regular) noun phrase to be coreferential with the same magnet or a with the other token of that object I was able to systematically quantify the coreferentiality requirements of both types of noun phrases. Definite noun phrases establishing or referring to a unique referent in the domain of discourse were predicted to prompt repetitive selection of the old noun token. Weak definites, which are hypothesized not to presuppose uniqueness, should prompt selection of the old noun token in fewer cases than regular definites.

In reviewing the videos, I counted a response as a new token selection if the participant either moved the new token magnet and repositioned it adjacent to the person magnet or repositioned the person magnet adjacent to the new token magnet. Critically, if the person magnet ended up next to the new token magnet, this was counted as a new token selection. Likewise, I counted a response as an old token selection if the participant moved either the old token magnet or the person magnet so as to be adjacent to one another. An example of something coded as a non-unique, or weak interpretation, of is shown in Figure 2.2, where each person magnet has been positioned with a different bus.

In 73 percent of the trials with weak definite NPs, participants selected the new token as the final referent, compared with 44 percent of the trials with regular definite NPs. An arcsine transform was performed to normalize the data as proportions by subject. Statistical analyses revealed that this difference was significant ($t = -3.3453$, $p < .0001$). Results are shown in Figure 6.1 (all error bars are standard error).

In addition to a traditional t-test, a mixed-effects linear regression model was performed so that subjects and items could be examined as random effects, representing the fact that both participants and stimuli are representative of a larger population. The analysis examines whether noun type could predict whether subjects would choose the previously selected token or the new token. This model effectively demonstrates what was shown in the t-test, which is that subjects were more likely to select the new token after a weak definite noun ($\beta = 1.9, SE = .39, p < .0001$).
I set up the study in order to examine how important listeners found it to map a definite NP to one particular place or object. Canonically, definites should correspond with a unique referent, and this turned out to be true in Experiment One for the regular definite NPs. However, subjects did not prioritize consistently mapping one object magnet to a weak definite NP. In 73 percent of weak definite trials, subjects chose a new entity to enact the story and opted not to choose the common-grounded (or “familiar”) entity. These data demonstrate that comprehenders interpret weak definite noun phrases as being significantly less referentially linked to a particular entity than regular definites are. Participants chose new noun tokens significantly more often after a repetition of a weak definite phrase than a regular definite phrase. This indicates that weak definites either do not establish a unique entity in the common ground or do not presuppose reference to a unique entity in the common ground.

Initially, it may seem surprising that the new token was also selected for regular definites 44 percent of the time, even though regular definites conventionally refer to a uniquely identifiable discourse entity. However, recall that the experimental displays used color to implicitly establish visually prominent (distinct) domains. In the post-experimental briefing of participants, I found that this manipulation was enormously “successful”: despite the fact that several fillers explicitly required the movement of items across visual domains, many participants still reported that they felt hesitant to do so. This eagerness to remain within the original color domain when completing the task may initially seem like an alternative explanation for the strength of these results. Participants’ willingness to choose two distinct tokens of the noun might be because the color domains coerce a type of possessive interpretation of each definite NP. That is, the newspaper is interpreted as something like the newspaper in the green area, and weak referentiality need not be a factor. However, the large difference in how readily subjects made this choice with the weak nouns versus the regular nouns suggests that even if possessive definite interpretations are a factor under these experimental conditions, there is a strong difference in the uniqueness presuppositions of weak versus regular definite NPs. The claim that weak definite NP do not necessarily establish or refer uniquely to a referent in
the way that conventional definite NPs do is consistent with these data, even given
the startling strength of my referential domain manipulation.

The surprising strength of the behavioral data alleviates all concerns that vi-

sual domain manipulation would be unable to provide an environment where weak
definite interpretations could be examined experimentally. This renders our eye-
movement manipulation redundant, and so those data were not analyzed in detail.
Figure 2.2  Experiment One in Action: Two still photograph frames from Experiment One, showing a non-unique interpretation of the phrase “rode the bus.”
Figure 2.3  Experiment One Results: Proportion of new token magnet selection when participants heard a regular definite NP or a weak definite NP.
3 Weak Definites and the Introduction of Discourse Referents

The results of Experiment One are consistent with claims that weak definites either fail to establish a novel referent or fail to presuppose referential uniqueness. Listeners’ frequent selection of a new object for repeated mentions of a weak definite NP indicates that the repetition did not presuppose uniqueness or evoke familiarity—either that, or the first mention failed to establish a discourse entity that would be seen as unique or familiar later. The kind of anaphora that might be expected, where comprehenders would select the same object for repeated mentions of the definite NP, only manifested in the lab with typical definites. This difference between weak and regular definites is substantial, but the heart of this difference is unclear. Perhaps weak definites do not effectively establish new entities in the common ground, thus leaving the listener with nothing familiar to refer back to upon the second mention. The other possible explanation, though, is that weak definites do not require this type of uniqueness, so a repetition of a weak definite noun phrase does not prompt the listener to search for a familiar object.

If this is the case, and my results are due to weak definites’ lack of uniqueness presuppositions, then these noun phrases potentially behave as *indefinites*. Indefinites are capable of establishing new referents into the discourse model, and contrast with informationally definite noun phrases, which require at least weak familiarity. The results of Experiment One do not rule out the possibility that weak definites
are merely indefinites in disguise. Experiment Two directly compares weak definites to both regular definites and indefinites, in order to assess the relative extent to which each can establish new discourse entities in the minds of comprehenders. Additionally, Experiment Two will reassess the lack of uniqueness in weak definites in a setting where uniqueness and referent-establishment can be teased apart experimentally.

A scene showing multiple matching discourse referents (e.g. a picture of two bicycles) should, without further distinguishing information, make sentences that use a definite NP (e.g. “the bicycle”) infelicitous, due to a conflict between the article and the visual scene. Likewise, if a noun phrase establishes a discourse referent (e.g. “a bicycle”), then a picture depicting multiple tokens matching that noun should also be somewhat awkward. However, a singular noun phrase that fails to establish a discourse referent may be an acceptable match for a picture with two like objects. In Experiment Two, I compare weak definites to indefinites and regular definites by using this method.

### 3.1 Experiment Two

#### 3.1.1 Methods

**Participants**

Thirty-three members of the University of Rochester community participated in this study for pay. All were native speakers of American English with normal or corrected-to-normal vision.

**Design and Materials**

Experiment Two is a scene verification task, where participants see text paired with pictures and are asked to rate the goodness-of-fit of text and picture. This study used a $2 \times 2 \times 2$ design, crossing number of critical token objects shown in the picture
(one or two) with type of article in the text (definite or indefinite) and type of noun (regular or weak). Written stimuli such as (2) and (2) below contain either a weak or regular noun paired with either the definite or indefinite article. The conjoined-NP form of the sentences was chosen in order to emphasize the simultaneity of the described activities, so that participants would not be confused by seeing both individuals depicted in the same scene. Pictures crucially show either one or two tokens of the critical noun (e.g. bike or bus). Twenty-four critical items were intermixed with 24 fillers. Examples of critical items are shown in Fig. 3.1.

(1)  Dean and Anne rode a/the bike.

(2)  Jerome and Sandy rode a/the bus.

Fillers were constructed by creating sentence/scene pairs for which there was either a very good fit or a very poor fit. These items allowed us to assess whether subjects were attending to the task. For example, subjects might see a picture of a man eating ice cream paired with a sentence such as “John ate pizza.” Failure to give this type of item a low rating, for instance, can be used as a built-in comprehension task used to exclude any subjects who are not attending to the experimental task. Additionally, fillers of this sort help mask the experimental manipulation, which in this case might otherwise seem obvious.
In the one-token picture condition, all noun-article combinations should be rated highly; that is, it is perfectly felicitous for “a bike”, “the bike”, “a bus”, or “the bus” to be used in a sentence referring to a visual scene in which there is only one bike, or only one bus present. It is in the two token condition where I expect differences to arise. For regular definites, I anticipate that participants will rate these scenes less highly; it is not felicitous for “Dean and Anne rode the bike” to be used to describe a scene in which Dean and Anne are each riding their own bicycle, since in such a scene the definite NP cannot refer uniquely. Since indefinites canonically serve to introduce a discourse referent, we might expect that indefinite text paired with two-token scenes would be rated less highly, due to the mismatch between number of referents (two) and number of new referents that the noun phrase is attempting to establish (one). The way subjects rate weak definites is of interest in this study: if they are rated on par with regular definites and indefinites, this would suggest that they do serve to establish discourse referents, and they may be essentially the same as indefinite noun phrases. If, however, weak definites are rated highly with two-token scenes, and this differs from indefinite ratings, weak definites are not simply odd-looking indefinites.

Procedure

Scenes and sentences were presented simultaneously on a computer monitor. Participants were instructed to rate the appropriateness of the written descriptions for each picture on a seven-point scale by typing the numbers ‘1’ through ‘7’ on a keyboard, with seven being the best fit, and one being the worst fit. Participants were given three practice trials during which they could ask questions of the experimenter before beginning the actual experiment. No feedback was given during the task.

3.1.2 Results and Discussion

Two participants were excluded from analysis for reversing the seven-point scale, as determined by their responses to filler items, which indicated the highest ratings for
semantic mismatches and the lowest ratings for strong semantic matches. For the remaining 31 subjects, I examined ratings for the critical picture/text pairs. If the critical NP establishes a new discourse entity or uniquely refers to familiar entity, two-referent scenes should be rated as less appropriate compared to one-referent scenes, since the number of common-ground entities will not match the number of referents in the scene. If, however, the NP does not serve to establish a discourse entity, this numerical mismatch of referents will not be infelicitous, and participants will accept two- as well as one-referent scenes.

As expected, when only one token of the critical noun was pictured, all text-picture pairs were rated as being highly appropriate (with averages of 6 or higher), as shown in Table 3.1. Ratings for the two-referent scenes were lower overall, as demonstrated in Table 3.2, and this difference in ratings between one-token and two-token conditions manifests itself as a significant main effect of Number of Pictures, as shown in Table 3.3.

<table>
<thead>
<tr>
<th></th>
<th>Definite</th>
<th>Indefinite</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regular</td>
<td>6.0</td>
<td>6.0</td>
</tr>
<tr>
<td>Weak</td>
<td>6.1</td>
<td>6.1</td>
</tr>
</tbody>
</table>

Table 3.1 Mean Ratings for One-Referent Scenes

<table>
<thead>
<tr>
<th></th>
<th>Definite</th>
<th>Indefinite</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regular</td>
<td>3.7</td>
<td>4.2</td>
</tr>
<tr>
<td>Weak</td>
<td>4.8</td>
<td>4.7</td>
</tr>
</tbody>
</table>

Table 3.2 Mean Ratings for Two-Referent Scenes

Thus I focus my analysis on the trials containing two-referent pictures, where I expect any effects of noun type or article type to appear. Recall that these scenes should be least acceptable when the critical noun phrase refers to or establishes a single entity in the discourse. The regular definites (e.g. rode the bike) show the lowest ratings, with an average acceptability of 3.7 on a seven-point scale. Participants did not like seeing sentences like “Dean and Anne rode the bike” paired with a picture of two people riding separate bikes. Regular indefinites (e.g. rode a bike)
Table 3.3  Three-way ANOVA for Experiment Two: F-values, p-values, and Significance indicators for Noun Type, Article Type, Number of Pictures, and corresponding interactions

<table>
<thead>
<tr>
<th>Factor</th>
<th>F-value</th>
<th>p &lt;</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noun Type</td>
<td>0.85</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>Article</td>
<td>1.0</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>Number of Pictures</td>
<td>47.1</td>
<td>0.0001</td>
<td>***</td>
</tr>
<tr>
<td>Noun:Article</td>
<td>3.0</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>Noun:Pictures</td>
<td>13.9</td>
<td>0.001</td>
<td>***</td>
</tr>
<tr>
<td>Article:Pictures</td>
<td>1.7</td>
<td>0.2</td>
<td></td>
</tr>
<tr>
<td>Noun:Article:Pictures</td>
<td>2.7</td>
<td>0.2</td>
<td></td>
</tr>
</tbody>
</table>

were somewhat better with these pictures, and averaged 4.2. Weak indefinites (e.g. rode a bus) were still better, at 4.7. But the weak definite NPs (e.g. rode the bus) had the highest average acceptability rating at 4.8. Mean ratings for two-referent scenes are shown in Table 3.2.

Qualitatively, the rank order of these data seem to fit our intuitions about which noun phrases are the most familiar or established in the common ground: regular definites are the worst match for multiple consistent objects, and they are the phrases that should involve reference to a familiar entity. Regular indefinites, which are capable of establishing new discourse referents, are somewhat better. Weak indefinites and especially weak definites provide the best fit for the two-object scenes, which indicates that they may not have the discourse-referential properties of either regular definites or indefinites.

A quantitative approach via an Analysis of Variance for the two-referent conditions showed a statistically significant interaction (see Table 3.4): article type matters more for the interpretation of regular nouns, or more specifically, for nouns that do not allow weak interpretations ($p < .05$). Two-referent scenes are generally more acceptable with nouns allowing a weak reading, whether the article is definite or indefinite. Regular definites are less acceptable with two-token scenes than regular indefinites are; however, the opposite is true for weak nouns, where definites are, if anything, better (albeit not significantly) than indefinites for describing
two-referent scenes.

Additionally, a mixed-effects linear regression model was used to predict the rating given to the description-picture pair, with article type and noun type as fixed effects, and subject and item as random effects, and found that participants were significantly more likely to give a two-token scene a high rating when it was paired with an indefinite article ($\beta = 0.72, SE = 0.058, t = 12.4$).

Participants were also significantly more likely to give a two-referent scene a high rating when the critical sentence in the display contained a weak noun ($\beta = 0.72, SE = 0.058, t = 12.4$).

There was also a significant interaction between article type and noun type in this analysis ($\beta = -0.72, SE = 0.086, t = -8.37$).

<table>
<thead>
<tr>
<th>Factor</th>
<th>F-value</th>
<th>p</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noun Type</td>
<td>6.5</td>
<td>0.02</td>
<td>*</td>
</tr>
<tr>
<td>Article</td>
<td>0.06</td>
<td>0.9</td>
<td></td>
</tr>
<tr>
<td>Noun:Article</td>
<td>4.2</td>
<td>0.05</td>
<td>*</td>
</tr>
</tbody>
</table>

Table 3.4 Two-way ANOVA for Two-Referent Conditions

Notably, weak indefinites are more acceptable with two referents than are regular indefinites. That is, definiteness does not seem to matter as much when the noun is one that can convey a weak reading. We speculate that this could be due to the lexical components of weak definites: even when paired with an indefinite article, these nouns, in the right context, might be preferentially interpreted as evoking an event rather than an individual. That is, in the sentence “Jerome and Sandy rode a bus”, *a bus* may evoke the activity of *bus-riding* rather than an individual discourse entity *bus*, in much the same way that weak definites (e.g. *the bus*) appear to do.

In order to address our question about whether weak definites are canonical indefinites masquerading with a definite article, we conducted a planned t-test between the weak definites and the regular indefinites, which shows these conditions to be significantly different ($t = -2.249$, $p$-value = 0.025), giving support to the claim that weak definites are not simply regular indefinites in disguise. Because the rela-
relationship between weak definite NPs and their indefinite counterparts is not known, we chose to explicitly compare the weak definite cases to the standard indefinite cases, where intuition and theory are more explicit.

These results suggest similar findings to a recent study with Dutch weak definites. In this study, Scholten and Aguilar-Guevara (2010) present subjects with sentences that contain a weak definite, bare singular (e.g. go to bed), or a typical indefinite noun phrase. Then in a follow-up sentence, which refers to the same item, subjects must choose whether to use a pronoun (e.g. it) or an anaphoric definite noun phrase. Their results showed that pronominal reference was much more frequently used when a typical noun had originally been introduced with an indefinite description than when a bare or a weak definite was presented. The authors conclusions with the Dutch anaphora study are similar to mine: namely that weak definites do not typically establish a discourse referent.

Figure 3.2  Experiment Two Results: Two-Referent Scene Ratings
4 Semantic Enrichment

4.1 Experiment Three-A

Experiments One and Two are consistent with claims that weak definites neither refer uniquely nor establish a new discourse referent. These results set weak definites apart from both canonical definites and indefinites. If weak definites are not impacting comprehension as other noun phrases with articles do, then this raises questions about their communicative contribution. If weak definites do not introduce entities into the common ground, and they do not presuppose that an entity is already a part of that common ground, then what information do they provide the language user?

One hint of their contribution comes from Experiment One, where participants most often selected the new token of a noun upon repetition when that noun was a weak definite. This suggests that comprehenders interpret weak definites as referring to a familiar or known type of event or activity, rather than to a known discourse referent. In other words, if both Trish and Amos “went to the movies,” subjects cared that Trish and Amos each engaged in a movie-going event, and not necessarily that they both went to the same film or theater.

The key observation here is that weak definites seem to convey an enriched meaning consistent with the typical event involving the noun. Going to “the movies,” for instance, means more than just being present at a movie theater—it has to do with the experience of watching a motion picture in public. An electrician called out to fix equipment at the cinema would probably not be described as having “gone to
the movies,” which typically involves buying a ticket, sitting in a comfortable chair, and watching a new release on a very large screen. This incongruity between the enriched weak definite, “go to the movies,” an an event that does not involve a typical movie-going experience is demonstrated in Example (1). The most colloquial weak definite NPs, which I avoid using in my experimental items because of large regional variations, appear to have very strong levels of enrichment; they also seem to resist strong reference. This is demonstrated in Examples (2) and (3), respectively.

(1) *Today the electrician went to the movies. She got called to fix a problem with one of the projectors.

(2) *Today the plumber went to the slammer. He had to fix a pipe that burst in cell block four.

(3) *Today Barry was out in the boonies. They have a really great new restaurant there.

The enrichment conveyed by the weak definite seems strong enough to clash with any contradictory information. This could help explain the unique informative role of weak definites as compared with regular definites and indefinites. Enrichment could be a major part of what weak definite NPs contribute to discourse. Experiment Three-A explicitly tests this hypothesis by measuring the strength of the enriched interpretations of weak definite NPs.

The examples above provide us with an interesting way to examine the relative strength of weak definite enrichment. By selecting semantic agents that have strong prototypical role associations, such as “electrician,” and pairing those agents with weak definites that have conflicting enriched activity interpretations, it may become possible to measure the relative strength of semantic enrichment. In essence, Experiment Three-A was created to assess what happens when an electrician goes to the movies—does she go to fix a projector, or does she go to see a film?
4.1.1 Methods

Participants

Nineteen members of the University of Rochester community participated in the experiment for pay. All were native speakers of American English. Three participants had to be excluded due to computer malfunctions during the experiment.

Design and materials

Experiment Three-A compared weak definites and regular definites in an interview-style task, in which participants were asked about their interpretations of sentences like the one in Example (4).

(4) The FedEx driver had to go to the farm/hospital.

Twelve experimental items were constructed by pairing a randomly selected agent from a list of agents with well-known prototypical activities with a randomly selected destination-type weak definite and a regular definite counterpart. Fifteen filler items were constructed using other agents associated with prototypical activities: these nouns were in sentential contexts that made it unclear whether or not the agent was actually engaging in the relevant prototypical activity.

Among weak definites, there appear to be two main categories of enrichment involved: some (e.g. play the piano, read the newspaper, walk the dog) seem to convey a habitual-activity reading, while others, particularly the ones denoting destinations (e.g. go to the store, go to the hospital, go to the doctor) express enrichment related to the typical activity that occurs at that location. I chose to restrict our efforts to this latter set of weak definites because scenarios where people visit a destination for a non-canonical reason (such as going to the store to do something

1These matched regular definites were only matched intuitively, by considering properties such as imageability and clarity of what types of activities are likely to take place at such a place. Future work may wish to address this more directly by collecting some kind of norming data on the place nouns ahead of time.
other than shopping) are more plausible than similar scenarios for habitual activity weak definites (such as reading the newspaper for a reason other than to learn the news). Well-known agents were employed in all the sentences, so that the typicality of semantic roles was maximally salient (e.g. mailman, cab driver, pizza guy), as in Example (4).

In Example (4), when the destination is a regular definite (e.g. the farm), listeners might typically infer that the FedEx driver was making a delivery to the farm (since, after all, making deliveries is what FedEx drivers often do). But if, as suspected, weak definites express a certain amount of event-typicality of their own, when the driver’s destination is a weak definite NP (such as the hospital), comprehenders might instead infer that the FedEx driver had a medical problem and was not making a delivery. In this way, weak definites might influence what language users understand such sentences to represent: their assumptions based on the agent might be overridden by the enriched meaning of weak definites, which could in turn affect the mental representation of a described event.

A targeted question was constructed to follow each item. The question for critical items directly addressed the typicality of the agent’s role. For example, after the sentence in Example (4), subjects heard the question in Example (5). This question was always a yes/no question designed to query the participant about the role of the agent rather than the enriched meaning of the weak definite so as not to bias participants toward constructing an enriched meaning if one did not exist organically. Filler questions addressed various components of the sentence so that participants did not begin focusing disproportionately on the sentence agents. A full list of items and questions appears in Appendix C.

(5) Was the FedEx driver making a delivery in the scene you imagined?

Procedure

The experiment was set up with the participant seated at one computer, and a research assistant, who served as the interviewer, seated at another computer. The
participant and the RA were able to see and talk to each other, but could not see each other’s computer screens. Participants were instructed to read the sentences presented to them on their computer screen and visualize the scene that the sentence described. The RA then asked the participant the critical yes/no question, targeted at establishing whether the participant had imagined a scene in which the agent of the sentence was engaging in their prototypical agentive role. The RA recorded each yes/no answer as it was given. Then the RA asked the participant to verbally describe the details of the scene they had imagined. This description was recorded as an audio file and later transcribed by the same RA. An example of how a trial unfolded is given in Example (6):

(6) ON SCREEN: *The FedEx driver had to go to the farm/hospital.*
RA: *Was the FedEx driver making a delivery in the scene you imagined?*
SUBJECT RESPONDS: *yes/no*
RA: *Describe the scene you imagined.*
SUBJECT RESPONDS

This set-up allowed me not only to explore whether the use of a weak definite as the destination would result in an enriched reading that “cancelled out” the typical inference about the activities of the agent, but also to obtain information about what details the enriched meanings, if there were any, actually involve.

### 4.1.2 Results and Discussion

Each yes/no response to the target question was recorded at the time of the study. These examples were counted for each condition so that an analysis could be done to assess whether responses differed meaningfully depending on whether the noun was weak or regular. “Yes” responses indicate that the subject imagined the sentence’s agent performing his or her typical agentive role. Since these roles were not compatible with an enriched meaning of the destination noun, a “yes” response suggests that the strength of the agentive role was greater than the enrichment of
the destination weak definite. “No” responses indicated that the participant did not envision the agent acting in a typical role, perhaps because the enriched associations of the destination noun were stronger than the agentive typicality; thus this response is expected to be more likely with weak than regular definites, if weak definites convey a greater degree of enriched meaning as hypothesized.

A mixed effects logistic regression model was used to predict whether a participant would answer “yes” using NP type as a fixed effect and subject and item as random effects. This model revealed that the difference in response frequency was significant ($\beta = -1.4$, SE = 0.4, $p < 0.001$). This indicates that the manipulation of noun type (weak or regular) had an impact on the way participants imagined these scenarios unfolding. The numbers of “yes” and “no” responses by condition, where “no” indicates that the participant did not imagine that the agent was engaging in a prototypical activity, is given in Table 4.1:

<table>
<thead>
<tr>
<th></th>
<th>“No”</th>
<th>“Yes”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weak Definites</td>
<td>72</td>
<td>30</td>
</tr>
<tr>
<td>Regular Definites</td>
<td>41</td>
<td>61</td>
</tr>
</tbody>
</table>

Table 4.1  Experiment Three-A: Yes/No Responses

As predicted, I demonstrate a much larger portion of “no” responses for weak definites than for regular definites. Specifically, after reading a sentence containing a weak definite, participants were more likely to say that the agent was not performing their prototypical activity in the scene the imagined, and after reading a sentence containing a regular definite, they were more likely to say that the agent was performing their prototypical activity. This pattern is consistent with claims that weak definites evoke a canonical activity type, and that this enrichment is robust enough to override event biases conveyed by the sentence’s agent.

Qualitative responses that were collected in audio recordings and transcribed by an RA, such as the one in Example (6), demonstrate that participants answering “no” nearly always imagined a scenario involving details about activities typically taking place at the critical destination.
In order to better understand these qualitative responses and their relationship to the interpretation of enriched weak definite noun phrases, I instructed an undergraduate assistant, who was naive about the critical manipulation, to annotate the subjects’ descriptions of their imagined scenes. The descriptions were annotated for three criteria: explicit reference to an agent’s accessory item (e.g. a FedEx driver’s package), explicit naming of the critical destination (e.g. farm or hospital), and explicit mention of the action typical of the agentive role (e.g. making a delivery).

All three of these criteria were hypothesized to be predictive of a “yes” response, where the typical agentive role was imagined. Accessories were thought to be predictive of a “yes” response, since the accessories have a strong association to the agentive role. For example, a FedEx driver who goes to the hospital to make a delivery must have had some sort of package with him, while a FedEx driver who goes to the hospital to get medical treatment need not have a package with him at that time. Explicit naming of the critical destination was thought to be predictive of a “yes” response, because in these cases, the particular identity of the destination may be more relevant or salient to the comprehender than in cases where the destination’s own prototypical role dominates. That is, when a FedEx driver delivers a package to the hospital, the prepositional phrase including the critical noun is a critical part of the imagined scenario; a FedEx driver who goes to the hospital because he is injured can be described as “going to get treatment” with no explicit mention of the destination. Finally, explicit mention of the typical agentive action was hypothesized to predict a response inclusive of that action, for obvious reasons.

A separate linear mixed effects regression model was performed for each of these three criteria with subjects and items as random effects. Mention of accessory was found to significantly predict “yes” responses ($\beta = 3.99$, SE = .50, $p < .0001$). Mention of destination was also found to significantly predict “yes” responses ($\beta = -4.48$, SE = .65, $p < .0001$). Finally, explicit mention of the typical agentive role predicted “yes” responses ($\beta = 6.05$, SE = .72, $p < .0001$). These results strongly suggest that the more details about the agent or the agent’s typical role that subjects included in their qualitative descriptions, the more likely the subject was to imagine
This closer look at the qualitative responses subjects made during the imagination task reveals that the mental representation built upon typicality can be quite detailed. Examining comprehension data in this way can offer strong clues about how language users both understand and envision scenarios that they learn about through linguistic report.

4.2 Experiment Three-B

The results of Experiment Two suggested that weak referentiality extends beyond unusual definite NPs: I found that weak nouns in indefinite phrases were rated as somewhat more acceptable pairs for two-referent scenes than were regular indefinite NPs. Recall that many defining properties of weak definites are linked directly to the noun, and not the article. Due to the lexical nature of weak definite noun phrases, nouns that appear in weak definite NPs may also convey additional enrichment when paired with the indefinite article. Experiment Three-B examines the interpretations of “weak indefinites” compared to regular indefinites, using the same method as I used in Experiment Three-A. This allows the role that the noun plays in weak interpretations to be directly examined.

4.2.1 Methods

Participants

Sixteen members of the University of Rochester community participated in the experiment for pay. All were native speakers of American English.

Design and Materials

The design of Experiment Three-B was identical to that of Three-A. Materials were also identical, except that the article in the critical noun phrase was indefinite, rather
Results and Discussion

As in Experiment Three-B, each yes/no response to the target question was recorded at the time of the study. “Yes” responses indicate that the subject imagined the sentence’s agent performing his or her typical agentive role. “No” responses indicated that the participant did not envision the agent acting in a typical role.

As in Experiment Three-A, a mixed effects logistic regression model was used to predict whether participants would answer “yes, using noun type as a fixed effect and subjects and items as random effects, but this model found only a marginal effect of noun type ($\beta = -1.05$, SE = 0.5, $p < 0.06$). The numbers of “yes” and “no” responses by condition, where “no” indicates that the participant did not imagine that the agent was engaging in a prototypical activity, is given in Table 4.2:

<table>
<thead>
<tr>
<th>Condition</th>
<th>“No”</th>
<th>“Yes”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weak Definites</td>
<td>68</td>
<td>28</td>
</tr>
<tr>
<td>Regular Definites</td>
<td>47</td>
<td>49</td>
</tr>
</tbody>
</table>

*Table 4.2  Experiment Three-B: Yes/No Responses*

At first glance, it appears that weak indefinites have less of a distinctive impact on the way participants translate utterances to imagined mental scenes than their weak definite counterparts do. However, when the results of Experiments Three-A and Three-B are directly compared in a model with article type (either definite or indefinite) as an additional fixed effect, no significant interaction between noun type and article type emerges.

These data demonstrate the same basic type of effect with weak indefinites as with the weak definites in Experiment Three-A: in both cases, weak noun phrases show stronger enrichment than comparable regular noun phrases. This lends support to claims that weak reference is largely attributable to the noun. However, this
result alone does not distinguish between claims that these nouns are anomalous exception words and my own hypothesis, that certain conceptual features of nouns allow for weak reference\(^2\).

\(^2\)Although all of the destination nouns used in this study were quite common, it is possible that they differed with respect to their familiarity. That is, farm may be less familiar as a concept than hospital, and this difference, if systematic across items, may cause the weak nouns to appear more enriched than regular nouns. Schulpen (2010) suggest a follow-up study to address this by comparing weak definites to modified weak definites, e.g. the new hospital, which does not allow a weak interpretation, but is matched for conceptual familiarity by virtue of being the same noun.
5 Categories and Semantic Features

5.1 Experiment Four

Experiment Three-A demonstrates that weak definite noun phrases frequently express meanings that are enriched by conventional knowledge of certain activities. Yet similar enhancement was found in Experiment Three-B, where these critical nouns were paired with the indefinite article. Experiment Three-B suggests that the lexical semantics of the noun contributes significantly to enriched interpretations that weakly referential phrases convey. This raises questions about why some nouns allow weak reference and carry an additional layer of meaning, while others do not.

To examine what these nouns have in common, we can look both within and across languages. The types of nouns demonstrating these properties largely fall into one of several categories, and this is true when looking at English alone, but more compellingly, the pattern holds across languages that have weak definites. This indicates that conceptual properties may play a larger role in the determination of weak reference than previously acknowledged. The following categories and nouns provide an incomplete list of the words found in weakly referential phrases:

- Mass transportation: bus, train, subway, ferry
- Mass communication: radio, news, newspaper, calendar, phone, TV (Irish), telly (British English)
• Service destinations: hospital, doctor, slammer, store, bank, bathroom, jail
  (Dutch)

• Outings: movies, cinema, beach, mountains

• Undesirable places: boonies, boondocks, ghetto, sticks

• Chore/routine nouns: dog, piano, dishes, trash, lawn, calendar

• Body and building parts: knee, eye, corner, window, stairs, wall, hall, doorway

Cutting across these categories is a notable interchangeability of token in contexts where weak interpretations are found: the enriched meanings of weak definites typically do not rely on token specificity. When one needs medical treatment, for example, any hospital is likely to offer the same set of services and thus the unique identifiability of the hospital is secondary. The same is true for things such as mass transportation, where the regularity of the route is more critical than the particular vehicle, which is interchangeable. If a person takes the bus to work every day, for example, she probably takes the same route, with the actual vehicle being inconsequential.

That weak definites fall into conceptual categories (across languages!) suggests that they might not be mere exception words in the lexicon. In Experiment Four I probe the conceptual origins of weak definite noun phrases by manipulating some of the traits that seem common among these category members. Specifically, this study investigates whether novel words representing novel concepts can license weak reference when the regularity of context and the interchangeability of referential token are manipulated. If the manipulation of critical semantic features can result in the addition of new weak-definite-capable nouns to the lexicon, we will be one step closer to understanding the conceptual underpinnings of these nouns and the mental representation of language more generally.
5.1.1 Methods

Participants

Eight adult native speakers of American English from the Rochester community volunteered to participate in the study for payment.

Design and Materials

Artificial lexicons can be a useful psycholinguistic tool, because they allow the researcher to create a tiny language that experimental participants have never been exposed to; any biases coming from frequency or co-occurrence in natural language can be bypassed by presenting subjects with a totally novel set of words (Magnuson et al., 2003). Because of the small set of weak definites in American English, and because of questions about how weak definites might originate in natural language, I chose to explore the conceptual properties of weak definites within the context of an artificial lexicon. To this end, I wrote a short science fiction narrative that was, except for the novel words, written in English. The chosen genre of the story was science fiction, because I felt that this would best allow a plausible context for both novel words and unfamiliar concepts. Unfamiliar concepts are necessary so that any artificial words are not merely translations of English nouns, which are potentially encoded for their ability to convey weak referential interpretations. All novel words were purportedly Martian lexical items that referred to alien concepts and were embedded in a story about life on the planet Mars. The Martian backdrop of the short story has the advantage of supporting a context in which novel concepts are likely to appear.

Two types of novel concepts were introduced in the body of the story: weak-valenced and regular-valenced. Weak-valenced concepts highlighted semantic properties associated cross-linguistically with weak reference, especially regularity of context and purpose, and interchangeability of referential token. For example, the concept “vlerter” was introduced as a special closet-type room where humans living on Mars routinely must enter to be hosed down with a special protective skin coating
to compensate for lack of atmospheric protection from UV rays. This is the exclusive purpose of vlerters, which are widely dispersed among both residential and commercial buildings in Mars City and can be used by anyone. Regular-valenced concepts, on the other hand, appeared in a variety of descriptive contexts, and individuality of the referent was emphasized by the content of the story. For example, the concept “ven” was introduced as a comfort device that can provide a range of palliative capabilities, including heat, ice, softness, or hangover cure, depending on the needs of the user.

At no point was either type of noun introduced in a context that forced a weak interpretation. Uniqueness was plausible in all contexts where a novel word was paired with the definite article. This means that participants critically were not trained to accept weak reference through exposure to phrases that necessitate non-uniqueness. Measuring subjects’ acceptance of weak reference with these nouns after exposure will then give me an indication of a) whether new weak definites can be created in the lexicon, and b) whether my hypothesized conceptual features lend themselves to weak definite interpretations.

Procedure

After giving consent, participants were informed that they would be reading a short science fiction story, and that they would be asked to answer comprehension questions both during and after they completed the story. They were also told that the story would introduce unfamiliar words, which would be underlined the first time they appear in the text. The manuscript itself was a printed document with four brief comprehension quizzes dispersed throughout. The purpose of these embedded comprehension questions was to encourage participants to read the material closely.

After completing the story, participants immediately performed the critical linguistic judgment task, where they rated the acceptability of forced weak interpretations in anaphoric expressions on a scale from 1 to 7. An example of what subjects rated is shown in (1), and the full list can be found in Appendix D.
Example (1) demonstrates that the final sentence, in the given context, must be a weak definite in order to be an acceptable summary of the prior scenario.

After they completed the critical task, subjects then performed a semantic rating task to evaluate the valence of our novel words. This involved subjects viewing statements constructed to target the conceptual features manipulated in the experiment and rating those statements for appropriateness on a scale from 1 to 7. These statements were of the type in Example (2) and (3).

(2) Vens are a personal item: everyone’s is special to them.

(3) Vlerters are ubiquitous in Martian buildings, and all serve the same function.

In other words, I wanted to see whether my novel weak concepts truly possessed the intended qualities of interchangeability, etc. These ratings were presented after the critical task so that they did not influence participants’ judgments about weak reference.

The final task addressed the concern that our novel words were simply translations of English words for existing “Earth” concepts: participants were asked to rate their confidence that they have learned and understood each novel word and then write a one-word definition of each novel word they learned – a task that should be difficult and provoke a variety of responses if my artificial lexicon successfully describes novel concepts.
5.1.2 Results and Discussion

After reading the story, subjects rated scenarios like Example (1) on a 1 to 7 scale for how well the final sentence summarized the paragraph. The paragraphs all contextually imply two distinct tokens of the critical noun. In Example (1) this is conveyed by suggesting that Ash and Padraig used vlerers in two distinct locations. The summary sentence contains a definite NP, the critical word, and then a VP anaphora. These VP-anaphora sentences are intuitively acceptable with weak definite NPs, so they provide a useful test environment for weak definite interpretations of my novel words. If the word allows a weak definite reading, I expect overall higher ratings on the scale, indicating that these VP-anaphora sentences provide a good summary of the prior scenario. On the other hand, low acceptability ratings would indicate that the novel word does not allow a weak definite interpretation.

The interest of this result critically depends on whether I managed to associate the conceptual properties of interest with the novel words introduced in the story. For this reason, subjects answered questions that were crafted to assess the perceived valence of each new word. The valence manipulation overall proved successful, with the mean values of ratings aligning with our intended semantic properties. A linear mixed model with subject and novel word as random effects and valence as the fixed effect revealed that the novel words overall conveyed the conceptual traits I was most interested in manipulating ($\beta = 3.01$, $SE = .50$, $p < .0001$).

An examination of each item’s rating, however, revealed that one noun was rated by subjects as not having its intended (weak) valence. This was “vurrit,” a type of vehicle, which some subjects conceptualized as a mass-transit vehicle (which emphasizes routine and interchangeability) and others as an individually operated vehicle (which emphasizes uniqueness and variability of use). This disparity of conceptualization among the subjects was revealed by my final task, in which participants were asked to define each new word they had learned. Answers for “vurrit” included (but were not limited to) “tandem bicycle,” “car thing,” “bus,” and “pull-cart vehicle.” In the valence assessment question for this lexical item, the mean rating was 4.6 out of 7, where the intended sense of the word was weak (the
low-end of the seven-point scale). Because the intended valence was not created, and more importantly, because subjects were split on the valence of this word, data were analyzed with “vurrit” excluded.

A t-test on the sentence ratings revealed a significant difference in linguistic judgment of weakly referential sentences: with a mean of 3.86, regular-valence nouns were rated as significantly less acceptable in weakly referential sentences than my novel weak-valence nouns, which had a mean rating of 5.03 ($t = 3.08, p = .003$). Additionally, a mixed effects linear regression model to predict rating was performed, with noun type as a fixed effect and subject and item as random effects. This model also supported that participants were significantly more likely to give a weak-valanced word a high rating in a forced weakly referential sentence ($\beta = 1.17, SE = .33, t$-value = 3.496). This demonstrates that the manipulations present in the narrative conveyed the semantic traits I intended, and that these differences in valence predicted the acceptability of sentences containing the novel nouns in a context that forced a weak, non-unique interpretation of a definite NP. Nouns that were introduced as having greater interchangeability and regularity of purpose were more likely to license weakly referential NPs.

As anticipated, subjects reported difficulty composing a one-word definition of each novel lexical item. For no novel word did all subjects report identical definitions; for some words, each subject reported a unique definition. This alleviates concerns that this task is merely assessing weak definite interpretation indirectly through a one-to-one correspondence between artificial words and existing English words (which may be weak). In order to further demonstrate that English translation did not mediate our subjects’ linguistic judgments, a linear mixed model was performed to assess whether the “weakness” of the English translation predicted the valence of the novel word. In other words, each subject’s translations were coded for whether the English word used has a weak definite interpretation or not. That is, if a subject translated “filochut” as “subway,” which can be used in weak definite phrases in English, this was annotated in the data set. Subject and word were included as random effects in the model. Whether English translation words
themselves allow for weak readings did not predict valence of the novel words (β = -.15, SE = .45, p = .74). Furthermore, English translation words’ weak-definite availability did not directly predict linguistic judgment of the critical weak definite sentences (β = .84, SE = .49, p = .90).

What I have shown here is that new nouns allowing non-unique definite NP interpretations are not simply grammatical exceptions to definiteness that are fixed within the lexicon, but in fact can be created under specific conceptual conditions. The less relevant a particular token is to an event, the more acceptable non-uniqueness or weak reference becomes. For example, “vlerters” are rooms or closets that people visit to take prophylactic measures against the sun’s harmful rays. All vlerters serve this and very little other purpose. In any building, at any time, any Martian can use a vlerter. The “ven,” by contrast, is a type of security blanket or teddy bear that provides a range of comforts depending on the owner’s present condition. Since a ven is owned by an individual and in a variety of ways under different circumstances, one person’s ven might not be so easy to replace with another’s. This lack of interchangeability makes weak reference for a word like ven relatively unacceptable to language users.

The data from this study help unravel the origins of weak definite noun phrases. Within the limited capacity of a short story that exposed readers to a set of novel words, key properties of those words were enough to license weak definite noun phrases. Although the conceptual categories and the cross-linguistic similarity of weak definites suggests that their presence in the English lexicon is not simply accidental, this study provides the first experimental evidence that these structures are non-artifactual in the mental lexicon: the conceptual properties of the lexeme, as well as the context, can license weak reference. Weak reference is a trait that corresponds to a comprehender’s mental representation of an event, as communicated through language.

This evidence, when combined with the results of Experiment Three-B, which shows that weak readings appear with indefinite articles as well as in definite contexts, leaves unresolved the contribution that the definite article makes in these
special NPs. While the non-canonical use of the definite article is what initially draws the attention of researchers to weak definite NPs, my experimental evidence largely illuminates the role lexical semantics and sentential context play in weak interpretation. The role of the determiner itself, while a hallmark of these phrases in conventional usage, remains unclear and is not informed by the current results.
6 Conclusion

6.1 Experimental Summary

These studies have examined a special case of definite noun phrases, weak definites. These phrases appear to allow non-unique interpretations of the definite article in certain governing and contextual environments. For this reason, they raise questions about the intersection of definiteness, language processing, and mental representation.

In Experiment One, I demonstrated that this special class of nouns indeed allows non-unique interpretations in definite noun phrases. In a task where participants reenact a spoken narrative, uses of the definite article elicit much less repeat reference to a known entity for nouns that have been observed to possess a weak reading. When hearing repetitions of weak definites, subjects do not preferentially pick a referent and stick with it, the way they do with anaphoric uses of conventional definite NPs. Weak definite noun phrases are indeed only weakly referential, even when comprehenders are mapping language to their visual environment through an action task. Listeners repeatedly select the same noun type, but only select the identical token referent in a minority of cases.

The data of Experiment One are consistent with a view that weak definites are merely indefinites in disguise. To examine this hypothesis explicitly, I designed Experiment Two to address the difference between weak definites and indefinites. This study revealed that weak definites neither create nor refer to an established discourse
referent, which sets them apart from both indefinites and definites, respectively. In a task where subjects rated how well certain sentences matched pictures, weak definite sentences were rated as high matches for a picture, regardless of the number of matching token entities that were depicted. This was significantly different than canonical singular indefinites, which were rated as being less ideal matches when two token items were depicted. Number of referents matters for indefinites, presumably because they establish new discourse entities, but number does not affect weak definite ratings, because weak definites do not serve to establish discourse entities.

Since weak definites behave unlike either canonical definites or indefinites, I sought to examine what unique contributions they might make to discourse. Experiment Three-A investigated one specific influence weak definites are claimed to have on mental representation, semantic enrichment. This study confirmed that weak definites evoke an enriched event-type representation to a greater degree than regular definites. In an imagination task, subjects were more likely to report that the agent of a fictional narrative (e.g. “taxi driver”) was not performing his or her typical role when the destination present in the narrative was a weak definite (e.g. “the hospital”) as opposed to a conventional definite noun phrase (e.g. “the farm”). In other words, the enriched meaning of a weak definite was strong enough to overcome assumptions based on agentive roles. By using this imagination task, I revealed a strength of weak definites’ contribution to discourse information not previously known. Experiment Three-B probed this question further with indefinite phrases with both weak and regular nouns and found similar results. This indicates that the weak noun primarily may be driving the enrichment effect, with minimal assistance from the article.

Experiment Four uncovered more concretely that weakly referential nouns are not a closed or arbitrary class of words by demonstrating that new weak definites can be created in the mental lexicon. This task manipulated a science fiction story, which allowed me to create a novel lexicon that was embedded in English grammar but retained unique conceptual properties and was not simply a translation of
English words and Earthly concepts. Weak interpretations correlated with broader conceptual properties of interchangeability and contextual invariability, and novel words introduced with those properties licensed weak reference in comprehension tasks. This suggests that conceptual properties may play a role in linguistic form. Weak definites in particular are shown to be attributable to certain semantic features, and are not simply linguistic artifacts or anomalies.

Taken together, the evidence I have collected suggests that weak definite noun phrases exist which do not presuppose uniqueness of the referent, and which do not establish a new referent into the discourse. They diverge from conventional definites and indefinites in those ways. They also make unique contributions to the discourse by conveying a strong degree of semantic enrichment, and in this way they may be a type of cognitive and conversational shortcut when a typical or conventional activity is under discussion. New lexical items that have consistent and strong typicality, especially where the activity involving the noun is more critical than the token or entity itself, can be adopted as weakly referential noun phrases as well; these conceptual traits are consistent both within and across languages.

### 6.2 Conceptual Knowledge and Weak Nouns

Weak definites are only as informative as our knowledge of the relevant concepts is complete. A phrase such as “go to the hospital,” in other words, becomes enriched insofar as the language users have an understanding of what a typical hospital-going experience is like. A naturally occurring example reflecting this enrichment is shown in Figure 6.1, a screen shot of a Facebook user’s status. Experiment Four demonstrated that the consistency of this conceptual knowledge is related to how acceptable weak interpretations are to the comprehender. While Experiment Four only offers a first-pass examination of what conceptual properties allow for weak reference, it provides a proof-of-concept demonstration that weak definites are not merely exceptional items within the lexicon. Rather, they are an open class whose membership is largely dependent on contextual invariance and consistency of con-
This explains why weak definites tend to fall into one of a handful of categories, such as chores (e.g. “take out the trash”), service destinations (e.g. “go to the bank”), media (“listen to the radio”), and public transportation (“take the bus”). This is true both within and across languages, which demonstrates that weak definites are not lexical anomalies or idioms. These categories have a notable number of concepts that readily appear in culturally common contexts where the particular identity of the referent is not critical. I was able to explicitly manipulate this in Experiment Four with novel nouns, although it remains to be seen whether novel categories could be created that support weak reference as well.

We can deconstruct a single example where two similar words differ in their interchangeability and regularity. Take “car,” which does not have a weak interpretation, and “bus,” which does. Both are ubiquitous and widely accessible land vehicles, but cars are not interchangeable in the way public buses are. My car is a unique make, model, and color; it does not have a set route along which it operates; it is not accessible to the general public, etc. Buses, on the other hand, operate along set paths that are not dependent on the particular vehicle; they are not individual possessions; they can be accessed by anyone for a small fee; and their route is more crucial than their physical form. Thus, it is not surprising that “take the bus” has a weak interpretation while “take the car” does not. The weak reading of “take the bus” may not actually refer, and so an act of “taking the bus” can felicitously describe a scenario that involves switching buses several times in order to follow a
particular route.

It is also possible that forced weak definite interpretations are a signal to the comprehender that there is regularity associated with the event. A colleague suggested giving subjects novel words in contexts where a weak interpretation is forced and then examining how this affects the subjects’ understanding of the novel word’s conceptual properties.

Another seemingly exceptional use of definite descriptions also hinges on conceptual knowledge and event interpretations. Take the examples in (1) and (2).

1. A Fox 40 whistle was sold on eBay.
The price was great.
#The sound was great.

2. A Fox 40 whistle was blown to start the competition.
#The price was great.
The sound was great.

Bosch (2007) argues that the use of the definite article in some continuations may be better than others based on the function (and related knowledge) of the concepts described, such as sound and price. It is possible that when a physical event is being described in Example (2), a physical property like sound is seen as a more likely continuation by comprehenders, or is more salient in the common ground (Cieschinger, 2008). Interlocutors must know something about whistles in order to understand that they can be bought and sold, and that they make noise; and event-based knowledge also seems to affect how felicitous different definite descriptions seem in continuations.

6.3 Unresolved Questions about Articles

Although I have developed laboratory methods that have revealed critical features of weakly referential definite noun phrases, several questions are still unanswered.
First, the exact role that the article and the noun each play in these phrases is still not clear. This class has largely been described as a class of *nouns* that have special properties, although the phrases of interest are *definite* NPs. It has been noted that synonyms of weak-definite-capable nouns do not always allow for weak interpretations themselves. This initially gives weight to the view that weak definites are primarily an exceptional (and possibly closed) class of nouns; but Experiment Four reveals critical conceptual features that license weak reference, and often these non-weak synonyms lack such properties. This suggests that weak definites are not mere lexical anomalies, and knowing this, one might then scrutinize the definite article and wonder what type of informational signal it is providing. It is, of course, the presence of the definite article in these phrases that raised linguists’ eyebrows in the first place.

However, the precise contribution of the article to these phrases is still not well understood. Experiments Three A and B showed comparable levels of semantic enrichment with these nouns in the right contexts, whether the article was definite or indefinite. Although it is the definite article that makes these phrases stick out as unconventional, it seems as though the nouns themselves license weak referentiality. Nor are those nouns anomalies or artifacts, which indicates that it is a set of conceptual traits corresponding to these nouns that licenses weak reference. So what exactly does the definite article do, and why is this a difficult problem to solve?

One relevant observation that may explain why the definite article is not critical to weak definiteness is that the link between definiteness and article use is heavily dependent on context. Experiment One showed just one way in which referential domain is flexible: when I painted each side of a magnetic board a different color, subjects tended to treat those areas as distinct worlds. Even though conventional definite NPs presuppose uniqueness, the dimension of uniqueness, or the relevant referential domain, may not be immediately obvious. For instance, Example (3) uses a definite article in a canonical way, but many scenarios exist in which this particular definite article fails to provide much information to a comprehender. In a context with only one skater, and with no other non-unique entities present, “the”
could theoretically disambiguate the object of “hit” right away. However, in a case where there are many skaters wearing pink fishnets, we might not be able to unambiguously select the correct referent until the very last word, “helmet,” which is separated from the definite article by six intervening words. Thus, articles in general may be able to help constrain possible referents to some degree, but may vary too much contextually to be strong cues for the formation of expectations.

(3) *Lois hit the skater who was wearing a pink helmet.*

In a recent set of studies, Klein, Li, Jaeger, Carlson, and Tanenhaus (in press) demonstrate that prenominal function words may help exclude impossible referents from consideration without necessarily prompting strong expectations concerning upcoming nouns. These studies examine the effects of Chinese classifier words on object selection in Mandarin. Chinese classifiers, like articles in English, appear in front of referential nouns, but the set of classifiers is much larger than the set of articles, and each noun can only appear with a particular classifier or classifiers. The transparency of the relationship between classifier and noun varies in the language, with some classifiers being a strong shape or functional match for a noun, while other classifier/noun pairings appear to be relatively arbitrary. Thus the classifier theoretically has the potential to constrain the set of possible nouns that could follow it, either via knowledge based on co-occurrence frequency or through conceptual goodness-of-fit. Klein et al. (in press) find in a visual world task that, while grammatically incompatible objects receive few looks after classifier onset, eye-movements to target items are not strongly facilitated by information from the classifier. For an abridged version of this paper, see Appendix E.

The Chinese classifier data are similar to a finding by Dahan et al. (2000), which involves a visual world comprehension task with gendered determiners in French. In these studies, where target and competitor items had different grammatical genders, comprehenders also looked away from grammatically excluded referents but were not typically faster to select a target based on disambiguating cues from the determiner.
Weak definites, which are identical in form to strongly definite NPs, may be one more case where the use of articles and related function words in context may vary too much to prompt strong expectations on the part of the comprehender.

Yet intuitions persist that the definite article may be meaningful in these phrases. The combination of evidence from these experiments suggests a possible explanation: weak definites appear in contexts where event-based invariance and token-based interchangeability is likely, and these phrases convey an enriched meaning with respect to the activity being described. Critically, this enriched meaning is one that is related to a speaker’s knowledge about conventional, familiar events. Thus the definite article may serve to mark familiarity—a role it has been argued to play—even in these exceptional cases.

In collaboration with Whitney Gegg-Harrison, Greg Carlson, and Michael Tanenhaus, new preliminary evidence has shown a significant difference between weak definite and weak indefinite descriptions. In these cases, it is the weak definites that convey stronger familiarity judgments. For example, subjects read a scenario such as the one in (4).

(4) Kent lives in coastal North Carolina with his parents. One of his favorite pastimes is collecting sea shells, and he picks them up whenever he has the chance. Twice a year, the family travel to coastal Florida to visit his grandparents. Kent wrote an email to his friend, saying: I went to the/a beach last weekend.

Then subjects were asked to rate on a five-point scale how likely Kent was to be in North Carolina versus Florida. Finally, they were asked, ‘If instead he had written, ‘I went to a/the beach last weekend,’ would he be MORE or LESS likely to be in Florida?’

After giving judgments based on the initial scenario, which contained a weak noun such as beach, subjects were then explicitly asked how their judgment would differ if the determiner had been switched, which allowed us to assess the unique contribution of the definite article when used with a weak noun in a weak context.
When the initial phrase presented was a weak definite, subjects were much more likely (85 percent of the time) to report that the switch to an indefinite would indicate that Kent was more likely to be in an unfamiliar location. When the initial phrase was an indefinite description, subjects only said that the switch to a definite would indicate a more novel location 35 percent of the time. These data suggest that the definite article may serve to mark familiarity in weakly referential phrases, even though much of the interpretation is determined by the noun.

6.4 Referentiality

A common thread here, and one that is illustrated in the experiments I have presented, is that weak definites seem to describe familiar activities or events, and not familiar or unique referents. Evidence from Scholten and Aguilar-Guevara (2010) with Dutch is similar to our own, and is consistent with claims that weak definites do not establish discourse referents.

There is a formal notion of linguistic reference and a processing-based psychological notion. In linguistics, reference describes the mapping between a label and the object that exemplifies it. Within psychology, this terminology is used more informally but often serves to indicate the relationship between linguistic form and mental representation rather than world-based denotation.

While weak definites are linked to mental representation, and thus may be psychologically “referential,” from a formal linguistics standpoint the lack of anaphoric support that weak interpretations offer suggests that they are non-referential. My experiments strongly suggest that weak definites do not linguistically refer.

This is not to say that weak definites are uninformative. On the contrary, I have demonstrated that weak definites are capable of evoking rich information about conventional event types. The power of expression in these special noun phrases is in their sense—specifically in their ability to convey typicality—and not in reference.

If the definite article serves to mark familiarity, these phrases may not be so exceptional, even for a linguistic analysis. Heim (1982), Roberts (2003), and oth-
ers have explicated analyses of definite descriptions where familiarity licenses the use of definite descriptions. Although Heim posits a definition of familiarity that requires a high degree of salience (e.g. prior mention, items appearing in the environment, etc.), Roberts allows for weak familiarity, which includes contextually entailed referents. Perhaps a further stretch in this direction may allow activity types, which are in the collective common ground, and which do not rely upon a unique referent, to license definite descriptions (i.e. weak definites) via familiarity\(^1\), and thus incorporate weak definites into analyses of definite descriptions more broadly.

\(^1\)For a thorough discussion of different analyses of definite descriptions, which is outside the scope of this dissertation, see Cieschinger (2008), who also addresses cases that seem to thwart conventional analyses.


A  Experiment 1 Materials

1) Jane is feeling silly today. Have her wear a hat.
   weak: She wants to impress her friends with how talented she is. Have Jane play the saxophone. Now have Bill play the saxophone too.
   regular: She wants to impress her friends with how talented she is. Have Jane play with the ball. Now have Bill play with the ball too.

2) Carrie hasn’t been doing so well. Give her a teddy bear.
   weak: Now take Carrie to the doctor. Take Robert to the doctor too.
   regular: Now take Carrie to the teacher. Take Robert to the teacher too.

3) Jason does chores to earn his allowance. Today, he brushed his teeth first.
   weak: Then Jason walked the dog. This evening, Beth walked the dog too.
   regular: Then Jason played with the cat. This evening, Beth played with the cat too.

4) Fiona likes to help out around the house. Make her fold her clothes.
   weak: Have Fiona take out the trash. Now have Craig take out the trash too.
   regular: Have Fiona wash the vegetables. Now have Craig wash the vegetables too.

5) Amos goes out a lot. This morning, he went to a cafe.
   weak: Later, Amos went to the movies. Tonight Trish went to the movies too.
   regular: Later, Amos went to the concert. Tonight Trish went to the concert too.

6) Raylene has the day off, so she wants to run around town. Take her to a museum.
   weak: Then take Raylene to the store. Now take Jerry to the store too.
regular: Then take Raylene to the tower. Now take Jerry to the tower too.

7) Martin used his weekend to run errands. First, he bought groceries.
weak: Then Martin went to the bank. After lunch, Jill went to the bank too.
regular: Then Martin went to the church. After lunch, Jill went to the church too.

8) Krista has a bad habit of letting things get out of hand before she gets help. Take her to her bus stop.
weak: Now take Krista to the hospital. Then take Jack to the hospital too.
regular: Now take Krista to the school. Then take Jack to the school too.

9) Ricky is notorious for disrupting any quiet environment. First he spilled a glass of milk.
weak: Then Ricky talked on the phone. Kara talked on the phone too.
regular: Then Ricky talked on the megaphone. Kara talked on the megaphone too.

10) Lana is obsessed with music. Have her buy a Beatles poster.
weak: Now let Lana listen to the radio. Let Cedric listen to the radio too.
regular: Now let Lana listen to the record. Let Cedric listen to the record too.

11) Rudy is a very literary guy. This morning, he wrote poetry in his diary.
weak: Later, Rudy read the newspaper. This afternoon, Patty read the newspaper too.
regular: Later, Rudy read the book. This afternoon, Patty read the book too.

12) Leo lives about a mile away from school. He grabbed his backpack on the way out the door.
weak: Then Leo rode the bus. Mia rode the bus today too.
regular: Then Leo rode the bike. Mia rode the bike today too.
B  Experiment 2 Materials

Regular:

Tim and Katrina went to a/the teacher.
Crystal and Nick read a/the book.
Elaine and Evan went to a/the church.
Jordan and Sam rode a/the lawnmower.
Zach and Celine went to a/the fountain.
Max and Maresa played a/the game.
Earl and Susan went to a/the museum.
Ryan and Sarah played with a/the cat.
Dean and Anne rode a/the bike.
Jed and Martha looked at a/the yard.
Austin and Patty took a/the car.
Thomas and Andrea sat in a/the chair.

Weak:

Martin and Diana read a/the newspaper.
Andrew and Whitney went to a/the doctor.
Terri and Stephan went to a/the bank.
Robert and Hannah sat in a/the corner.
Brett and Beth went to a/the beach.
Jerome and Cindy rode a/the bus.
Jacob and Kathy rode a/the train.
Keith and Kaitlyn played a/the piano.
Domenica and Gerry went to a/the store.
Benedict and Traci walked a/the dog.
Nick and Patricia mowed the lawn.
Amos and Trisha took the streetcar.
C  Experiment Three Materials

Experiment Three-A
Critical Scenarios and Target Questions (Regular Definite, Weak Definite, Question):

The UPS driver had to go to the church.
The UPS driver had to go to the store.
Was the UPS driver making a delivery in the scene you imagined?

The pizza delivery guy had to go to the lawyer.
The pizza delivery guy had to go to the doctor.
Was the pizza delivery guy delivering pizza in the scene you imagined?

The florist went to the monument.
The florist went to the beach.
Was the florist delivering flowers in the scene you imagined?

The mailman went to the courthouse.
The mailman went to the library.
Was the mailman delivering mail in the scene you imagined?

The cab driver had to go to the museum.
The cab driver had to go to the bank.
Did the cab driver have a passenger in the scene you imagined?
The FEDEX driver had to go to the farm.
The FEDEX driver had to go to the hospital.
Was the FEDEX driver making a delivery in the scene you imagined?

Experiment Three-B
Critical Scenarios and Target Questions (Regular Indefinite, Weak Indefinite, Question):

The UPS driver had to go to a church.
The UPS driver had to go to a store.
Was the UPS driver making a delivery in the scene you imagined?

The pizza delivery guy had to go to a lawyer.
The pizza delivery guy had to go to a doctor.
Was the pizza delivery guy delivering pizza in the scene you imagined?

The florist went to a monument.
The florist went to a beach.
Was the florist delivering flowers in the scene you imagined?

The mailman went to a courthouse.
The mailman went to a library.
Was the mailman delivering mail in the scene you imagined?

The cab driver had to go to a museum.
The cab driver had to go to a bank.
Did the cab driver have a passenger in the scene you imagined?

The FEDEX driver had to go to a farm.
The FEDEX driver had to go to a hospital.
Was the FEDEX driver making a delivery in the scene you imagined?
D  Experiment 4 Materials

D.1  Artificial Lexicon Science Fiction Story

I always managed to get myself to school in the mornings until I became a PhD student. Now the brightly colored vurrit number 47 picks me and two other students up almost every day. Now that 30’s the new 21, why not give us an embarrassing ride along with all that education? Seriously, the thing looks like a jet-propelled Radio Flyer wagon. If my field were sociology or intellectual history, or maybe even xenopsychology, my sellout book would be The Infantilization of the Intellectual. But that’s not my field.

Marty, our own personal soccer mom, just started driving this year (I actually hate to call it “driving,” since it’s more like “pushing” or “rolling”) and he hasn’t quite gotten the hang of how to maneuver the vurrit, so I always buckle up. It’s a good idea anyway with the low gravity, since one small bump can really send you flying. The shoulder straps and my protective outerwear make me, I’m sure, resemble a toddler in a snow suit, strapped into a Bugaboo, incapable of scratching my nose and on the brink of having to pee. All the jostling around doesn’t help. It also doesn’t help that I overslept, thanks to this being a chronin, and I almost broke my neck getting out of my hukkel in a hurry to meet Marty.

Special environmental conditions aside, I start my day the same way every graduate student does: by asking myself what the hell I’m doing here. Fortunately, I can crank the volume on my ossiphone to drown out my inner dialogue, which is sounding like a broken record even to myself lately. I listen to Irish language ossicasts on my commute, since Ireland just built a hood one kilometer down the Third Spoke. If you knew anything about MC real estate, you’d be raising your eyebrows. The
Third Spoke isn’t as developed as the First or Second Spokes, but it has 24 hour water and power, and filochut capsules jet pretty frequently from there to the City Center. The only other small country with as big a hood is Israel, which has basically spearheaded development of the entire Ninth Spoke. No surprise there. But this kind of a land-grab isn’t typical of the Irish.

New Connemara represents an enormous commitment to the preservation of the Irish language. For well over a century, the Irish government has distinguished districts in Ireland where Irish is the predominant language. But an influx of Polish immigrants, upper-middle class English speakers looking for vacation homes, American television shows, et cetera, have diluted these Gaeltacht areas. Since everything in MC is strictly regulated, the Coimisin na Gaeltachta can control the population and, to some extent, the linguistic exposure in New Connemara.

I can feel the air change as Marty pushes the clunky vurrit through the tunnel and into the campus bubble. Gates University is the most terrestrial hood besides the City Center. You can breathe freely throughout campus, and all indoor areas are gravitized to normal. The nicest part, though, is the grass. I never knew how much I’d miss grass. And while many of the hoods here have AstroTurf (the name turned out to be pretty appropriate), here at Gates, the Department of Agricultural Engineering has used our campus as a terrestrial lab space. We have a beautiful arboretum that even has a controlled squirrel population, but the grass is my favorite. At least once a week, free from the protective glipper required outside of hoods, I flop down on the grass in the main quad. There is never rain, no weather at all in fact aside from the delightful breeze from the air circulators, and the temperature on campus remains steady and comfortable year-round. If you can ignore the lack of smog and insects, it’s practically California.

I don’t have time to be homesick today, though. Marty dumped me off on the quad, and I walked straight to my office in Mason Hall, a square latex building (seriously), red like everything else here, with artificial brick-like molding that gives the building an authentic old Earth factory feel. “Greetings, Earthling!” said my officemate, Anssi.
“Doesn’t that ever get old?”

He looked at me over the rim of his glasses for a second and pretended to think about it. “Not really.”

“Fuck off,” I smiled. Anssi keeps me sane.

“I never use it on Katya, because Russians are a whole nother class of alien, as far as I’m concerned,” he mused.

“Agreed. You’re here early. Have you fed the gang?” I started peeling off my outermost layer of clothing, which was a thin, shiny glipper that I’d ordered in Revlon red. It looked a little like a mylar balloon, only the material thankfully didn’t crinkle when you walked.

“Just getting around to that. I made a new chart, this one with boxes for when we give these guys light.” He handed me a printed calendar grid with a glossary of special symbols at the bottom. Like all Gates students, we are required to share an office with a small animal population. Anssi, Katya, and I were assigned crawfish this year, and the adult female’s eggs had just hatched.

“Shouldn’t Livestock Engineering be giving you feeding charts?” I asked. I stood next to Anssi by the tank and peered in. Big Mama was hiding behind a giant chunk of shimmery areostike, and a few of the baby crawfish were clinging to a plant. The water in the tank had started to look like weak tea.

“Those guys are in over their heads monitoring all the populations. I thought I’d pick up some of their slack,” said Anssi. He opened the lid to the tank and sprinkled in some tiny food pellets and turned on the nutrition laser. Big Mama stuck her head out from behind the log to investigate.

“I don’t think they’re that backed up. Don’t tell me you plan on doing some experiments of your own in here.” Interaction with animals, while required for psychological as well as practical reasons, was also strictly regulated. Anssi looked at me and said nothing. “Anssi, I’m serious. You’ve told them Big Mama’s eggs hatched, right? Anssi?”

Silence.
“???, this is on you if something goes wrong. At least clean the tank before it starts to reek.”

“Are you kidding? If we wait a while longer, the water will be interchangeable with the department coffee. That’ll save us a few trips downstairs per day.”

[break for a set of comprehension questions]

“Dude, hey. How’d your meeting with Ash go?” Yuqing looked up at me from one of the comfy seats in the corner of the campus Starbucks. Yuqing was another student in my department, and we shared an advisor, Ash Everett.

“The usual,” I sighed and plopped down in the chair next to her. For the record, I don’t take plopping down for granted anymore: it can only really happen in gravitized enclosures. “I reminded him of what I’ve been working on for the past four years, he dodged all my questions about potential postdoc placements, and then he made some incredibly genius comment that will take me all month to interpret.”

“If he meets with me at all, I count that as a success.” Yuqing began writing a list of some sort on her wrist with a gel pen.

“??! You’re so full of it. You’d find away to count a publication in Cosmos as a failure,” I said, referring to the premiere interplanetary science journal.

“You’re probably right, dude, but I have his meals to compete with.” Our advisor’s got an appetite, and is a little bit of a space cadet, no pun intended. “I mean, how could I be worth his time? Bill Gates himself would have to have seen 2500 marks lying on the AstroTurf before it would have been worth his time to bend over and pick it up. With Ash, it’s more like 50 marks, but for us, every ment counts. Did you eat lunch? I just beasted two grit muffins. If I’d known you were coming, I’d have saved one.”

“Lunch with the Pod People? Nope,” I said, referring to our cliquish group of classmates, most of whom were MC locals. “But don’t worry, Anssi had some extra noodle salad so I ate that. He keeps that bottle of sriracha in our office, so it was pretty tasty.”

“You mean that spicy Asian crap with the rooster on the bottle? My dad used to love that shit.” She wrinkled her nose but didn’t look up.
“Sriracha, yeah. He bartered for it at the Chinese yeshag on the Second Spoke. It must have come off one of the recent grocery shuttles. I’m more of a Cholula or Tobasco girl, but I haven’t managed to find any around here.”

“These aliens don’t know a thing about spice. Still, it’s amazing how Anssi’s not part of the MC social amoeba,” she said as she pulled back her sleeve to continue writing her list of things to do. It was a good point. There was a small number of people my age who had grown up here, and Gates gave those students priority in admissions. The oldest native on record was, in fact, only three years older than I was. He’s not a grad student, though. He’s too busy enjoying his celebrity status and making money participating in experiments, like most of his cohort. MC was the only metropolis for tens of thousands of miles, so people’s attitudes could be very provincial.

“I don’t know. Anssi’s different.”

“Stella, what the hell are we doing here?” Yuqing had put down her pen.

“Do you mean Mars or do you mean grad school?” I asked.

I strapped in the vurrit for my ride back to my flat on the Eleventh Spoke and waved goodbye to Nelly, the night pusher. Mars City is laid out like a giant wheel, with the City Center in the middle. The City Center, like Gates, is climate controlled and gravitized to near-normal levels, although it doesn’t look as fancy or terrestrial. No grass. No AstroTurf even, except for in Highland Park, which is the only public sector of MC built to simulate seasons. Sounds fancy, but it just amounts to having an ice skating rink in the “winter” and an outdoor pool in the “summer.” Only the rich could afford to join the high-priced ski trips to Olympus Mons, the tallest mountain in the Solar System. Of course there wasn’t really snow on it, but the resorts there imported the stuff from the poles, and since it never melted, they could control the snow levels with virtual impunity. Twelve commuter filochut pipes protrude from the City Center along the Spokes infrastructure, and since private vehicles are expensive and restricted on Martian soil, most hoods have built up around the Spokes.

Gates, as well as most of the North American sponsored hoods, is on the Twelth
Spoke. I take the little vurrit to work instead of public transit like the filochut because I live two kilometers up the Eleventh Spoke, which is a hodge-podge of utility buildings, Eastern European apartment hoods, and university subsidized student housing. The undergrads are guaranteed two years of housing on campus in the dorms, but after that they typically join us graduate students in the Oldenborg hood in Upper Eleven. I could take an inbound 11 filochut capsule down to the City Center, then transfer to an outbound 12 capsule to get to campus, but Marty and Nelly push their family’s vurrit directly between Oldenborg and Gates twice an hour, six am to ten pm. (They earn a little extra cash from the University for their services, and it saves me loads of time. Even if it is a little embarrassing.)

Not that time works the same way here as it does back home. The time difference is something I’ll never get used to. Days here are longer by about half an hour, so every hour has 61 instead of 60 minutes. To cover the rest of the loss, we have a sort of time adjustment every two months called the chrondin where the clocks reset. I still wear the Earth watch my mother gave me when I graduated high school, which only confuses me more, but most clocks here are thankfully satellite-linked, so I don’t have to keep track of the time changes or chrondin schedules myself. The Pod People give me grief about my early bedtime. I never have adjusted to the longer days. It just means a little more time curled up in my hanging slumber hukkel.

Nobody locks their doors in this hood–there are more surveillance cameras than plasma screen TVs–so I walked into my flat and once again removed my outer layer of protective clothing. When I first arrived at Gates, the guys in my department told me that I should embrace low-gravity sports as a way of coping with the environmental change, and the girls suggested I enjoy adding a whole new layer of fashion to my wardrobe. Truth is, I don’t much like sports or shopping. They suck even more on Mars.

I don’t relish dressing up in an overpriced glipper just to look like a mylar balloon. Or playing baseball when every hit is a home run. Life in a space colony is not the progressive utopia you imagine. It’s also not a tech-savvy bastion of
erotically androgynous geniuses wearing sleek synthetic fibers. My apartment can be gravitized with the flip of a switch, but I’d kill for a gas range and a toilet that didn’t choke on everything. The reddish dust gets everywhere and screws with circuitry and ball bearings and sinuses. Your boogers are full of Mars at the end of the day. Also the sun is really dim, since Mars is 1.5 times farther away from the sun than Earth. Which means 2.25 times less light, or sunbathing in twilight. And all your clothes take on this rusty hue. Supplies and consumer goods are rationed, so everything in my apartment has a DIY look. More Gismo than Martha Stewart. We use a lot of Elmer’s liquid glass on Mars, and we wear Carhart’s new Stellar Endurance line. Think Teflon-coated carpenter jeans. Not cool. And we still haven’t managed to cosmify bowling. (On the other hand, I do have the most amazing sleep skein ever, called a ven. I got it from my friend Nav. A heavy weight, and yet effortlessly light, the ven comforts and consoles, almost like a meditation experience. Even if you’re drunk, sick, or both, my ven makes it all worthwhile, at least until the next morning. My ven could kick the ass of whatever favorite teddy bear, hot water bottle, or blanket you might have.)

Mister Biscotti started crawling up the bars of his cage the second I walked in the door. He’s my foster hamster, from one of the Livestock labs. This RA who works there has a crush on me, so he lets me keep Mister Biscotti here in my flat, although technically he’s university property. I take him out of his cage every night and let him run around the floor in this plastic pink ball. He loves running in his ball. I went to the cage and stuck a yogurt treat through the bars, which he promptly fit inside his cheek before scurrying into a toilet-paper roll. Biscotti and my ficus bonsai are the only living things I have to come home to at the end of the day. Not that they’re the only living things in my building. The upstairs neighbors screw a million times a week, and the girl across the hall cooks a lot of questionably smelling food, so there’s life here. I just don’t connect with much of it.

My yanawang remote receiving glove buzzed around my thumb. I watched an alpha-memo from Anssi unfold along my index finger: “Am watching the crawfish. There is no joy in my life.” I don’t have a great reply, so I crawl up the ladder and
into my hanging hukkel and go to sleep early, tangled up with my ven.

[break for a set of comprehension questions]

The next day on the vurrit ride I make some progress with Irish. The voice of the avuncular narrator reminds me that I should focus on what I can say, not on what I can’t. His voice makes me think of dew and grass and coral beaches, miles of peat bog covering layers of human existence. Earth, if not Ireland. And he tells me to listen, again, to the following dialogue. A woman asking a man whether he wants coffee. Yes, please. Milk? Sugar? No, thank you.

The best language learning tracks are repetitive like that. You hear a dialogue that describes some quotidian interaction, then you hear parts of it slowed down, repeated, over-pronounced, ad nauseam. When you’re so bored that you want the vurrit to explode with you in it just for fun, then you know you’ve learned something. So I contemplate suicide and smile when Seamas tells Una for the fifth time that he would not like sugar, but milk would be fine. My kind narrator, Fearghal, comes on at the end to offer more encouragement, and I think I actually say “Slan! to him out loud when he signs off, because Marty chuckles.

I de-mylar in the office and shove my glipper into a drawer. Anssi hasn’t arrived, so I take some time alone with the baby crawfish. At first glance, they don’t exist, but after I hold still in front of the tank for a minute or two, they shyly creep out from behind a hollowed-out log. They’re tiny, translucent, and crescent-shaped, like toenail clippings from an infant. I drop some shrimp pellets into the tank and check Tuesday AM off on the chart Anssi’s made. The giant pellets hurtling toward the areostike gravel were enough to startle most of the hatchlings back into their hukkels, which are suspended inside the hollow log.

It looks like Katya’s been here recently. Her workspace is so immaculate that I notice every slight change. This time, I can see that her computer monitor has been repositioned and her zen rock garden has been rearranged. Seriously, she’s got this tiny sandbox on the side of her desk that has a few round stones in it, along with a miniature rake. I think the gimmick is that you can meditate while you rake the sand. She also has a ven in her office chair. Which isn’t a bad idea, come to think of
it. My part of the office is a wreck, and I wouldn’t know the difference if a burglar had rummaged through my stuff looking for a few marks. (That’s what the currency here is called, by the way—the Mars Buck, or marks. It currently trades about three to the dollar.) Anssi’s space is the Hollywood version of mine. My desk is messy in a tedious way: unsorted journal articles, uncompelling graphs of pilot data, crusty coffee mugs, faded photographs from back home. His stuff has character, shows-and-not-tells that he is a paragon of post-exodus indie geekery. Cock sauce and peanut butter on the shelf next to antique copies of the Dungeon Masters Guide and Nabokov. Even a copy of the Bible in Tok Pisin. He has a laser etching pen next to his keyboard, and his screen saver is a scrolling display of the largest prime number ever discovered, which is over thirteen million digits long. Under his desk is a case with a spackler inside it. Spacklers are technically woodwind instruments, but since they are rooted in the ground, performance takes on a visual quality as well. The red Martian dust gets propelled up the main pipe and shoots out in various patterns, all depending on the notes you play. The spackler was the first instrument to be created on Mars, and it’s taken on a special folk status here. Only Anssi would manage to squeeze a spackler under his desk. If there’s no joy in his life, it isn’t because his office is drab.

The Department of Geopolitical Psychology has about thirty grad students at any given time, and eight tenure-track faculty members. It’s meaningless to count, though, since everyone at Gates collaborates with people in other departments. Everything and everyone in MC strives to be multi-purpose. If you can’t be both a doctor and a lawyer, then you’d better at least be intersex. As the name suggests, our department tows the line. Mars, the big red frontier, is a hotbed of geopolitical conflict, and many students have a background in international relations, environmental analysis, or xenopsychology. Ash’s lab is a little different: we’re linguists. Psycholinguists, really. MC isn’t just a physical land-grab. English took over Earth decades ago, mediated somewhat by the sheer mass of China, and English got an early start here as well. But many of the settlements along the MC spokes are heritage enclaves. Hebrew has an enormous presence, for example, in the Israeli sectors. Mandarin is ubiquitous, so it doesn’t hurt that I minored in Asian languages.
But my thesis work is on New Connemara, hence the Irish language podcasts.

New Connemara represents the largest extraplanetary settlement proportional to its Earth base, in terms of both size and population. And unlike English, Chinese, or even Hebrew, Irish was not exactly thriving back home. I’m doing longitudinal work that charts the growth of the language community and the bleed into other language groups here in MC. The thing is, I don’t believe the story that New Connemara is just a cosmic Gaeltacht, offering a space for native speakers within its confines. I suspect there’s an active language-spreading policy in place, and I want to find out the details and, more importantly, track its success. Yuqing works on a similar project with Chinese. Lately this involves her sitting with elderly Mandarins drinking tea and playing majiang. Sounds like a pretty good gig to me, but it drives her nuts. She says she left Earth to escape that scene. Fair enough.

“They’ve assigned you a priald,” Ash poked his head into my office and startled me.

“Sorry, what?”

“Padraig O’Cinneide. Will be your priald, when you begin your field work. Designated to show you around, answer any of your questions, and probably attempt to control what you see there in New Connemara.” Ash chipped away some paint on my doorframe with his thumbnail. Eye-contact was somewhat rare with my advisor. I guess he saves it for special occasions or something.

“The Irish don’t want me wandering around their compound alone or what?”

“Having a priald will make your job a little harder, but this shouldn’t come as a surprise. Nobody appreciates an academic. You’ll learn that soon enough.”

“I think I’m there, Ash.”

“Then I’m doing my job.”

[break for a set of comprehension questions]

“Dude, the SAS conference submission deadline is today. AND my ven is at the cleaner’s, so I’m sleeping like ass. I’m going crazy here,” said Yuqing. I could see
the faint shadows of yesterday’s skin list creeping out from under her sleeve. “Do me a favor and get me a misto?”

“Sure, whole or skim?” I set down my messenger bag in the chair next to hers and glanced at the line: short. We try to time our Starbucks breaks to minimize contact with the undergraduates.

“Man, skim. I’m too lazy to be anorexic, but I’m not looking to be a heifer.”

I returned a few minutes later with two mistos–heifer-style for me–and breathed a sigh of relief. I was in a happy place: full gravity, warm cup, friendly face. The Pod People are mostly caffeine free, so they’re unlikely to show up here. Coffee on the whole is an expensive and bitter Earth novelty to anyone who grew up here, and a priceless comfort to those of us who did not. “Here, drink up,” I said as I handed Yuqing her misto. Her face mirrored my own relief for a split second.

“Thank God, dude, I really need this. I have five hours until the deadline.” After savoring the first sip, she immediately put her game face back on.

“SAS is in Nanjing this year?” The Social Architecture of Speech conference is a big one for us, but like most academic conferences, it happens on Earth.

“Yeah, and I have to be there. Well, not BE there, but you know.” We had the option of sending a collaborator from a closer university or of delivering the talk ourselves via kinewang, the multi-sensory interplanetary network (6S internet) receiving glove, a fancier yanawang, which fits on your hand and allows you to impose a holographic cast, called a mianmian, of your presentation. Sending your mianmian to Earth via kinewang costs about as much as a plane ticket from New York to Los Angeles. And that’s if you own your own kinewang, or have access to one, which we thankfully do in Ash’s lab.

“Proxy or Hollywood-style?”

“Hollywood, def. I want the glory. Ash can pay for it. He’s so totally my intellectual sugar daddy. I’m writing up my stuff on the spread of Taiwanese lexical items among Sino-Martian teenagers.”

“Taiwanese words? Here on Mars?” The stalemate over the issue of Taiwanese independence continued back home, with Taiwan functionally-but-not-officially
self-governing, but only Mainland delegations had been allowed to engage in ex¬
errestrial colonization.

“Yeah, these kids go to great lengths to be rebellious. I guess they found some shit on the inter-internet. The Taiwanese word for ’big boobs’ or something. Seriously, kids, even I managed to find cooler ways to rebel. It’s a stupid side project, but maybe it will get accepted.”

“I wish my main squeeze were as cool as your side project.” I meant it.

“Oh man, I can’t have this fight with you today,” she rolled her eyes. “I have to write the abstract.”

“What fight?”

“The one where we argue about which of us is more academically screwed.”

“There is no fight, Yuqing, because that person is always me.”

“Always I. And at least Ash likes you. You guys, like, grok or some shit.”

“Always I sounds douche-noob, FYI. I do have lower expectations than you, and more experience with space cadets and foodies like Ash. But I am way in over my head with this Irish thing. They’ve assigned me a priald.”

“They don’t want you bumbling around the Gaeltacht unsupervised, huh? Afraid you will contaminate their brainwashing dens?”

“Muinphiars? They’re not brainwashing dens, they’re linguistic cleansing treat¬
mament centers,” I rolled my eyes. “If they’re threatened by me, I’m actually kind of flattered. I’ve got to get a hold of this Padraig guy, though, to figure out what kind of a leash he wants to keep me on.”

“Padraig O’Cinneade? He is your priald? He can put me on any kind of leash he likes,” she winked. She had set down her misto and pushed aside a notebook.

“How do you know him?”

“You know Vinotas?” She was referring to an annual graduate student soiree.

“I know what it is, but I’ve never been. I figured it would be a kind of pathetic meat market, where you could get a low-gravity headache, eat synth cheese
and crackers, and protect your friends from desperately hooking up with the first engineer that comes along. I’ve never understood why you go every year.”

“Yeah, Vinotas is totally that lame. I just know the one year I don’t go will be the year something amazing could have happened to me. Besides, there is one redeeming factor, dude, and that is the Padraig O’Cinneade factor. He’s def the most interesting thing to come out of the Neuroeconomics Department.”

“Is he a redhead? With a name like that, it seems likely.”

“You got a problem?”

“No, it’s a very precious mutation.” I had a problem.

“Don’t worry, when you meet Padraig, you’ll want to be a baby factory for alien mutants.”

I’m not sure which is more annoying: Padraig O’Cinneade himself, or his harem of jealous Moiras from the Third Spoke Gaeltacht. Good, sweet Fearghal from my ossicasts, who by the way is probably bald and not at all red-headed, did not prepare me to converse with a gaggle of teenage girls. And I tried greeting my chaperone in Irish, but he cut me off immediately and told me that, in his role as priald, he was happy to accommodate an Anglophone. The clenched mandible behind the smile indicated that he’d go to great lengths to avoid hearing me butcher his native tongue. So English it would be, except for when we were visiting any of the linguistically sterile muinphiars, where all young Gaeltacht citizens go for routine cognitive linguistic therapy. In muinphiars, Padraig told me, I would be allowed to take notes but not to speak at all. I resisted my urge to ask him whether my breathing might also disrupt the careful linguistic cleansing that happens in muinphiars, but then I remembered what Ash said about academics and how they’re never appreciated. I also knew full well that very few outsiders were allowed in the muinphiars at all. Ash sure could pull some strings.

The Coimisin na Gaeltachta has closed-circuit monitors on every corner of the hood, but that’s practically the norm in MC. What stuck out was that somehow the Gaeltacht had furnished itself with real grass and was almost as lush as the Gates
campus. “Do you have a real bog too?” I asked Padraig. He didn’t seem to hear me, but the Moiras following us snickered, and I buried myself in my notebook.

A very soft sort of music was playing over speakers mounted near the cameras. I listened carefully and barely made out the sonorant tones of what sounded like a Gregorian chant. Mental note to ask Anssi whether he knew anything about the folk music scene in New Connemara. As a pack, we made slow progress touring the grounds of the hood. Padraig led, I stumbled along in his wake, and the gaggle fluctuated in size behind us. Padraig met several acquaintances along the way and stopped to converse. In Irish, of course. After the first two or three such encounters, I gave up trying to understand what was being said. Despite his reluctance to talk to me, Padraig was a definite extrovert who was happy to talk to everyone else.

[break for a set of comprehension questions]

The best part about this new field assignment? The filochut commute. When I first got to MC, filo rides were awkward, because the pneumatics give you a gleeful sensation deep in your diaphragm, and most riders can’t help smiling a little. Kind of like infants who smile when they’re gassy. Shooting through the Spoke pipe network in a small capsule with a bunch of smiling Martians is more disconcerting than you might think. But now that I’m used to it, I really like that cabin pressure can trick my body into feeling happy for a few seconds. In that moment, I don’t just feel content, I feel elated to the core to be here.

Since I had to change filo capsules at the main station, I decided to walk around the City Center for a bit, and maybe grab some xiaolongbao for dinner. Like I said, the City Center’s no Gates, but it does have moderately decent gravitation, a Starbucks with free inter-internet, and an Ex-Urban Outlet, for all your overpriced glipper needs. There’s also a pretty vibrant night market scene. The biggest market is the one here in the City Center, but you can find yeschags all over the place. Strolling their streets is practically the number one leisure activity here on Mars. They each have their hood or Spoke specialties, but yeschags are all fundamentally the same: you can count on seeing hand-made crafts, pirated DVDs, spackler busking, knock-off clothing, and street food. An iron-oxide mineral dug from the
harsh Martian soil, called areostike, had replaced old earth crystals in the mysterious gypsy yeschag culture, so you could buy cheap areostike pendants or meditation crystals or whatever. There are plenty of ramalamadingdong hippies on Mars. But I go to yeschags for the street food, not the areostike. The Gates Student Handbook warns us not to eat at yeschags, since many of the vendors don’t have a food distribution license. You hear horror stories about people falling ill with the Martian runs, which is exactly what it sounds like, and other communicable diseases wrapped in your favorite contraband dumpling. But there’s no easier, cheaper, or more delicious place to pick up the latest Earth blockbuster and a quick bite. Most things are less than a mark, fifty ments a pop. Unlike the Pod People, many of whom were raised on vitamilk and nutrijerky, I require flavor to survive.

And the yeschags are about the only place where you can get flavor in MC. The City Center Yeschag has oyster omelets, daikon sausage, coffin sandwiches (bread stuffed with delish), pear milk, and even flavored condoms. All manner of crops and livestock are in very short supply here, as you might imagine. Shipments from Earth take between one and four years, depending on the current interplanetary distance, and are frankly too expensive to waste on food, which is considered a luxury and not a necessity among the locals. Attempts to farm and ranch on-planet are still in experimental stages, and it’s usually the fruits of these experiments that get used, illegally, in yeschag fare. But to have hot, juicy dumpling sauce drip down your face and onto your glipper? Don’t think I haven’t eyed my own crawfish once or twice as a potential snack. Not that I’d do that to Anssi, who is, like most Martians, a vegetarian. He uses more condiments than your average Pod Person, but his diet is still maybe the most alien thing about him.

I stashed a bit of pomelo and gave it to Mister Biscotti when I got back to the Eleventh Spoke hood. Mister Biscotti sniffed at my fingers, which probably still smelled like pop-hoppers. Then he crawled gingerly onto my palm and picked up the citrus in his front paws before shoving the treat into his cheek pouch and scurrying back into his cage. Just then, my yanawang thumb buzzed. I can’t afford mianmian projection or a kinewang glove for 6S casting, but no grad student could
survive without at least a yanawang for 2S connections. The 2S glove gives you all local (i.e. planetary) audio-visual access, and also has a biweekly data upload from Earth. Don’t even ask me how much I pay for service—it’s on par with rent, and you know renting an apartment on Mars ain’t cheap. The glove receiver buzzed again—probably that Padraig or a wangamarketer. I hit the little red “ignore” jewel on the tip of the yanawang pinky and considered hopping into my vlerter.

Depending on your outdoor exposure, and the current planetary-to-sun distance, every few days on Mars you need to cake yourself in a special compound that counters the effect of solar radiation. Every public establishment is required by Martian statute to have a vlerter, and private vlerters get inspected pretty regularly to make sure they’re functioning correctly. I stripped off my glipper and a few marks and ments fell out of the pocket in the inner lining. Fuck it, I’ll pick them up later. Then I peeled off my Calvin Kleins (how terrestrial of me) and stepped into the tiny closet. Traveling to and living on Mars is instant exposure therapy for any degree of claustrophobia you might have: you either get over it really quickly, or you go crazy. So the looming porous walls of the vlerter didn’t bother me, and neither did the loud pulsing noise of the prad shooting out at my skin. I know all grad students are a tad masochistic, but I kind of like the sting of the vlerter. The dry substance, which is called prad, has to be propelled with enough force to penetrate your pores, and then a special binding liquid comes out as a mist to seal it all in. The first few months I was here, I felt dirty all the time, like I was suffocating in prad, but now I’m used to it. Imagine having makeup all over your body. Or being airbrushed before, not after, your photo is taken. Some people get special vlerter compound, like chromo-prad or luci-prad, which gives them multicolored skin and things like that. Every nightclub in MC has at least two or three blacklight-sensitive people running around trashed. And you can actually get high just from licking them. Back on Earth, we used to dye patterns on our skin with flavored KoolAid, which seems lame by comparison. It didn’t prevent cancer, and we could never get a really even tone out of it, but it was something we just did, at sleepover parties or in the high school locker room.
I stepped out of the vlerter and threw on a robe, and then walked back out into
the living room and saw the message light blinking on the yanawang I’d tossed
aside. Must have been whoever buzzed earlier. I sat down, slid the yanawang onto
my hand, and hit the “play” jewel.

“Hi Stella,” said the unmistakable voice of Padraig O’Cinneade. “I wanted to
do my part as your priald and follow up with you at the end of our first day together.
I’m looking forward to working with you on your project. I’m about to start a 36-
hour muinphiar treatment, so I won’t be available via yanawang or kinewang. But
we’ll set up our next meeting after that. Oh–don’t forget that we have a chrondin
coming up. I know Earth folks have a hard time keeping track.” As if I’d ever cast
my mianmian to him. All in all, his message was strangely charming, even though
he managed to bring up both my other-planetness and my linguistic shortcomings
in such a brief message. But I hadn’t think about how speaking English to me was
going to contaminate him. A 36-hour muinphiar stay? The Gaeltacht was more
strict about linguistic exposure than I realized.

And then another message. Someone must have called while I was getting hosed
down by the vlerter. “Heya Stellz,” crooned the voice of my brother. His voice
always knocks the wind out of me. The Earth data dump must have happened while
I was getting dusted. The audio message would have been recorded by Laz earlier
in the week and then uploaded. “Shane got your package. And he’s just at a point
when he can open things with a little help from me and Lana. He’ll be the only kid
in day care with a real Martian glipper. Did I tell you those are getting trendy here?
I’d have you send one for Lana, but God knows they’ll be out of style by the time
it arrives. Shane talks about his Aunt Stella a little bit now, but before you let it
go to your head, you should know that he also babbles about the plumber, the gas
attendant, and even Miss Rose. You remember, that large soprano at church? She
kind of blows his mind. Anyway, I don’t have a lot to say. Mom still pretends you
might walk off the next UFO and need to be fed, but she’s really proud of her space
cadet. We’re all saving up to cast a family mianmian at Christmas.”

Most people think language research is about understanding what people say, or
what they write, or even just understanding language independent of humans. But one thing I’ve learned is that what people don’t say can be just as informative as what they do say. For instance, Laz had no message for me from Lana, even though I remember having thrown some areostike earrings in that parcel for her, and he didn’t say anything at all about Dad. In other words, there had been no progress back home on Earth: Lana still hated me, and Dad was still gambling

D.2 Comprehension Questions

The narrator is
a) waitress
b) a social worker
c) a graduate student
d) a Pod Person

The narrator sleeps in
a) a hukkel
b) an ossiphone
c) a vurrit
d) New Connemara

Marty operates a
a) areostike
b) vurrit
c) glipper
d) hukkel

The Third Spoke conveniently has frequent
a) power outages
b) dust storms  
c) glipper sales  
d) filochut capsules  

Protective outerwear is called  
a) a glipper  
b) a yanawang  
c) a chrondin  
d) a spackler  

What did Stellas friend Naz give her?  
a) areostike  
b) Teflon-covered jeans  
c) a ven  
d) Olympus Mons  

True or False:  
You can sleep with a chrondin.  
A yanawang fits on your hand.  
Life on Mars is utopian and sophisticated.  
The narrator dislikes shopping.  
There are many yeschags in Mars City.  
A hukkel is a piece of equipment used to cook food.  

Which of the following is NOT a feature of a glipper?  
a) It protects you from the elements.  
b) It makes you look like a mylar balloon.  
c) It fits on your hand like a glove.  
d) It can be taken off inside the Gates campus bubble.
The Gaeltacht is
a) a nickname for the fifth spoke
b) Gates Universitys student center
c) an Irish settlement in Mars City
d) Stellas name for her Martian classmates

Which of the following does Anssi NOT have in his office:
a) a hukkel
b) a spackler
c) sriracha
d) a copy of the Bible in Tok Pisin

Padraig OCinneade is Stellas
a) priald
b) boyfriend
c) mianmian
d) kinewang

Which of these items can be used to send a holographic image?
a) kinewang
b) areostike
c) chrondin
d) muinphiar

Yuqing is busy
a) making herself a misto
b) writing her thesis
c) submitting a conference abstract
d) learning how to play a spackler
D.3 Experimental Questions

D.3.1 Weak Definite Acceptability Ratings

Please read the following paragraphs and then rate how naturally the target sentence paraphrases or follows from the paragraph. You are judging how natural the target sentence sounds as a description of the events in the paragraph. 1 means “very unnatural and 7 means “very natural.

1) Yuqing got picked up this morning at her Sixth Spoke apartment by a vurrit driver named Harriet and rolled to work. Stella got picked up by Martys vehicle in front of her Oldenberg flat and rolled to work.
Yuqing rode the vurrit to work today, and so did Stella.
1 2 3 4 5 6 7

2) Anssis family ran out of the special protective anti-radiation dust. Coincidentally, Ashs lab also ran out today too. Ashs lab manager bought new prad for the lab from a catalog, while Anssis mother purchased some prad from a friend. The lab manager bought the prad, and so did Anssis mother.
1 2 3 4 5 6 7

3) Ash was outside all day long, and needed to do a radiation cleanse when he got home. In the Gaeltacht, Padraig also decided today was a good day to take protective measures from the suns rays. Ash used the vlerter, and so did Padraig.
1 2 3 4 5 6 7

4) Yuqing presented at a conference that took place on Earth by broadcasting a mianmian of her talk. That same day, Laz decided to splurge on a high-tech, live action Christmas message for Stella. Yuqing cast the mianmian, and so did Laz.
1 2 3 4 5 6 7

5) Katya wanted to pick up a few snacks at a market in the City Center. Over in the Ninth Spoke, a pack of Moiras felt like browsing the latest selection of pirated DVDs available in their neighborhood. Katya went to the yeschag, and so did the Moiras.
1 2 3 4 5 6 7
6) Ash needed to let his wife know that he would be home late from the office. Anssi wanted to make a date with a girl he met at Vinotas. Ash talked on the yanawang, and so did Anssi.

7) When Yuqing began doing research in the Mandarin hood, she was assigned a liaison. When Stella started to investigate the spread of Irish in Mars City, Padraig was in charge of showing her around. Yuqing was assigned the priald, and so was Stella.

8) Last month, Ashs daughter celebrated a birthday, but people got confused about the time change and ended up being late. Then a few weeks later, Stella overslept and missed an important meeting when the Martian calendar was being reset. Ashs daughters birthday was during the chrondin, and so was Stellas meeting.

9) Padraig heard too much English and sought treatment on campus. After one of the Moiras returned to the Gaeltacht from a week long camping trip with some friends, her parents sent her for linguistic treatment in the Hood. Padraig got sent to the muinphiar, and so did one of the Moiras.

10) Ash was invited to present at a conference in Amsterdam and wanted to cast a mianmian since he couldnt be there. Padraig got a talk accepted at a workshop in Limerick, and his department agreed to pay for him to cast a mianmian as well. Ash presented on the kinewang, and so did Padraig.

11) Katya decided it was time to clean the crawfish tank and all its contents, including the chunks of mineral. Meanwhile, Nav thought her favorite crystal gypsy earrings were looking a little dingy. Katya cleaned the areostike, and so did Nav.

12) Stella wanted to go from her flat in the Third Spoke to the City Center to buy new clothes. Anssi planned to go from campus to his aunts house in the Fifth Spoke hood. Stella took the filochut, and so did Anssi.
13) Ash ate lunch in a hurry and dropped food on his glipper. Stella spilled coffee on herself in Starbucks. Ash soiled the glipper, and so did Stella.

14) Yuqing got drunk and wanted something to help her sleep off her hangover in her hukkel. Anssi pulled a muscle and needed something to help soothe his pain while he sat in the office. Yuqing used the ven, and so did Anssi.

15) It was late on Saturday night, and many people in Oldenborg decided it was time to go to sleep. Stella climbed into the hukkel, and so did her neighbors.

16) Pod people love the performing arts, which play a big role in the annual Feast of Martius, where the Gates Campus Audio-Visual Orchestra performs one big routine. Both Anssi and Padraig performed in the routine. Anssi played the spackler, and so did Padraig.

### D.3.2 Valence Confirmation Questions

Please rate the following statements on a scale from 1 to 7. 1 meaning “not at all true and 7 meaning “very true. Vurrits can only be operated on predetermined, public routes.

1234567

Prad comes in many different types.

1234567

Vlerters are ubiquitous in Martian buildings, and all serve the same function.

1234567

Mianmians allow the viewer to see a holographic image in real time.

1234567

Yeschags are all pretty similar: if you’ve seen one, you’ve seen them all.
Yanawangs critically allow the user to communicate with anyone in Mars over the local network.

Prialds are assigned to serve specific, unique liaison functions.

Each calendar chrondin is observed for a different and special reason.

Someone needing linguistic treatment could go to any available muinphiar for care.

To talk to someone on Earth, any kinewang will do, if you have the money.

Areostike pieces all look almost identical.

Filochuts only operate along the Spoke lines in Mars City.

Glippers are interchangeable: they all look and fit the same.

Vens are a personal item: everyones is special to them.

Its a little weird to sleep in someone elses hukkel.

Spacklers are all played the same basic way.

D.3.3 Definition and Confidence Assessment

Next to each word below, please write the Earth/English word that seems most similar to you. Then rate, from 1-7, how well you feel you understand the meaning of the Martian word. 1 means not at all and 7 means very well.
E Incremental Interpretation of Chinese Classifiers and Massifiers

E.1 Background

This is an abridged version of a working manuscript I have written with collaborators Renjie Li, Florian Jaeger, Greg Carlson, and Mike Tanenhaus.

E.2 Introduction

Both mass and count nouns in Mandarin require a classifier when following a numeral, a universal quantifier, or a demonstrative, as in (1) and (2):

(1) Na BEN shu hen gui. That CL-volume book is expensive.

(2) Wo he le liang WAN tang. I drank two CL-bowl soup.

Which particular classifier is used for any given noun is grammatically rather than semantically determined, although the classifier may have an inherent meaning of its own and may evoke salient properties of the nouns with which it combines (Huang and Ahrens, 2003). Scholars have categorized classifiers and classifier/noun pairings in several ways. Huang and Ahrens (2003) distinguish classifiers by whether
they coerce individual, kind, or event meanings of the noun. Gao and Malt (2009) group classifiers by how evocative they are of the conceptual features of the nouns with which they pair. For example, they identify classifiers that evoke well-defined properties, prototype classifiers, and arbitrary classifier groups. Gao and Malt (2009)’s work raises the important observation that Chinese classifiers are unlikely to be uniformly informative about which noun might follow, with some being best viewed as generic or default (Myers, 2000).

In this paper, we will primarily focus on the Cheng and Sybesma (1998) taxonomy, which sorts classifiers based on ontology (see Krifka (1995) and Yang (2001) for similar divisions). The authors distinguish two main types of classifier use in Chinese: count classifiers, which combine with nouns to denote entities that appear individuated in the world, and mass classifiers or “massifiers”), which combine to denote substances or unindividuated groups. While the count classifiers have been of more interest to psychologists because of their potential relationship with categorization and conceptualization, massifiers are more analogous to English measure phrases, such as “three bouquets of lilies” or “that bottle of wine.” Although, unlike Gao and Malt (2009), Cheng and Sybesma (1998) posit a division that only distinguishes two types of classifier and ignores salient conceptual dimensions, it is not a cognitively arbitrary division: the authors assume a pre-linguistic ontological distinction between mass and count concepts.

Yet the mass/count distinction is not syntactically realized in Mandarin and all but a few exceptional nouns require classifiers following numerals or determiners. Given this, Chierchia (1998) and others have suggested that perhaps all nouns in Chinese are basically mass. Even if this is so, it does not follow that Chinese nouns cannot be conceptually distinguished as massifying and classifying. It is possible that a conceptual mass/count distinction may manifest itself as a difference in the way the noun meanings are understood by the comprehender. In this paper, we take an experimental approach to the following question: is there evidence that classifiers and massifiers exert a different influence on language processing? And if so, in what ways?
The following studies we report investigate the time-course and nature of classifier and massifier comprehension, and focus in particular on four questions:

- To what extent do English measure phrases influence language comprehension online?
- To what extent do classifiers in Mandarin Chinese mirror these effects?
- Do count and mass classifiers have an equivalent impact on language processing?
- Do these effects occur at a structurally superficial level, or do they reflect deeper syntactic and semantic processes?

### E.3 Experiment One: English Measure Phrases

We first sought to gain experimental traction in this domain by examining the online comprehension of English measure phrases. English has a grammatical mass/count distinction that leads to the use of measure phrases to denote mass referents. These measure phrases, as in (3) and (4), have a word order that is similar to equivalent Chinese phrases. Note that (3) refers to a grouped quantity of count referents, while (4) refers to a true substance.

(3) Harriet smoked two PACKS of cigarettes.

(4) Gretchen opened a BOTTLE of gin.

These measure words come from an open class and frequently can refer explicitly to a container item or a group array when examined as isolated words. Because of the strong intuitions that English measure phrases are semantically contentful, we might expect a prominent effect of the measure word on referent resolution.

In order to gain fine-grained temporal information on the comprehension of these measure phrases, we use the visual-world paradigm, which allows us to infer
participants’ understanding of our linguistic materials at any point in the speech stream by gauging where the subject was looking (Tanenhaus et al., 1995).

E.3.1 Methods

Participants

All 16 participants were native speakers of English from the University of Rochester community. They had normal or corrected-to-normal vision and reported normal hearing, and were paid for their participation in the experiment.

Materials

The experimental task involved selecting a target from a display of four groups of pictures on a computer screen, as shown in Figure E.1. A target group, its phonological competitor group, its classifier competitor group, and the classifier competitor’s phonological competitor group were always present in quadrants on the monitor. Fourteen critical instructions containing a measure word, such as “Choose a head of lettuce,” were given to subjects aurally along with a unique array of pictures. The targets were selected for imageability and familiarity, and were either substance or collective items. A second item that pairs with the same linguistic measure word was chosen as the classifier-consistent competitor. For example, the measure phrase “a head of” was paired with “lettuce” as a target and “broccoli” as a classifier competitor. For each of these nouns, an imageable phonological cohort was chosen, which shared initial phonology but could not typically be paired with the target measure phrase. “Letters” was matched with “lettuce” and “broth” was matched with “broccoli,” for instance.

Each group of pictures had three identical tokens of an object, in order to maximize the felicity of using a measure phrase for reference. That is, there were three heads of lettuce on the screen so that an instruction to “choose a head of lettuce” is neither an awkward use of the indefinite nor emphatically contrastive with “head
Figure E.1  English Measure Phrases Visual World Computer Display: Screen shot of English measure phrase “head of” referential selection trial, which shows four groups of pictures: lettuce, letters, broccoli, and broth, and a central fixation cross. Instructions in the general condition were “Choose a picture of lettuce.” Instructions in the specific condition were “Choose a head of lettuce.”

of broccoli.” Trial order and picture quadrant were randomized for each participant. A fixation cross was located in the middle of the screen to allow the eye-tracking software to recalibrate periodically throughout the experiment.

Instructions were recorded by a naive native speaker at two different levels of classifier specificity. A general, or uninformative-classifier instruction, such as “Choose a PICTURE of lettuce” was matched with a specific, or informative-classifier instruction like “Choose a HEAD of lettuce.” The only difference between the general and specific condition was the degree of informativity of the measure phrase: in the general condition, all pictures in the display were consistent with the classifier, and in the specific condition, only the target and the classifier competitor were consistent with the measure phrase. Half of the critical items subjects received had a general classifier and half had a specific classifier.
Fillers, which were randomly mixed with critical trials, referred to one of the phonological cohorts as a target, also half with a general and half with a specific instruction. This was to prevent subjects from being able to identify the target picture as being one that had both a phonetic and a classifier-consistent competitor on the screen concurrently.

**Procedure**

After giving informed consent, participants were fitted with a head-mounted eye-tracking device and headphones. After calibration to the subjects’ eyes, participants were told that they would hear instructions to click on certain pictures they would be shown. Three practice trials appeared, and unless the subject had any questions at that time, the experiment began immediately. Pre-recorded stimuli played over the headphones at a comfortable volume, and there was a short break halfway through the study to prevent ocular fatigue. The study took approximately thirty minutes, after which participants were debriefed and paid for participation.

**E.3.2 Predictions**

This study compares two levels of measure-word informativity, general and specific. In the general condition, all items in the display are consistent with the instruction up through the measure phrase, which provides no real information about which picture is the target. Comprehenders must wait for the noun to begin to select a target. In this condition, we expect to see standard phonological cohort effects, where participants will begin to look at the two similar-sounding groups shortly after noun onset. The target “lettuce” and its phonological cohort “letters” should compete for selection until a phonetic point of disambiguation (POD) can be processed, approximately 200 milliseconds after it appears in the auditory stimulus, which is the approximate time necessary to plan and launch a saccade.

We expect participants to behave differently in the specific instruction condition. Because “head of” is only consistent with “lettuce” and “broccoli,” looks to those
two items might rise soon after the classifier onset. Because the measure phrase is not consistent with the phonological cohort, “letters,” the usual cohort effect should be absent in this condition, if participants rapidly integrate measure-word information when understanding reference to mass items. If, however, measure words do not have a rapid impact on language comprehension, we may still see signs of phonological competition in the specific condition, or we may just not detect any increased attention toward the classifier-consistent competitors like “broccoli.”

E.3.3 Results

To examine the effects of measure phrases on sentence comprehension, we aligned all trials to the onset of the noun and then plotted the proportion of looks to the four object groups on the screen as the instruction is uttered. As predicted, we found evidence of phonological competition in the general condition, as shown by increased looks to both the target and its phonological cohort approximately 200 milliseconds after onset of the noun. While looks to the classifier competitor and its cohort pair begin to fall at this point, looks to the phonological competitor distinctly do not start falling until around 500 ms after noun onset, in other words, not until the phonological point of disambiguation is processed. Proportions of looks to the four item groups in the generic condition are shown in Figure E.2.

In the context of the more restrictive specific measure words, the pattern of looks is different. Just before the onset of the target noun, looks to the two classifier-consistent groups already attract approximately 70 percent of fixations. No noun has been mentioned, but information present in the measure word has impacted subjects’ expectations about what will be referred to in the instruction. This is shown in Figure E.3. The target diverges from the classifier competitor around 275 ms after noun onset, which is faster than in the general condition, and shortly after noun onset but before the phonological POD. Looks to target groups in the two conditions are plotted together in Figure E.4.
E.3.4 Discussion

The pattern of looks at the noun onset in the general condition corroborate the idea behind our stimuli: participants are looking at all four items about equally often. Our generic measure phrase was truly uninformative, as intended, and looks begin to diverge approximately 200 ms after the noun begins, which is very much in line with what we know about the time-course of planning and executing saccadic eye-movements. After this point, looks to the target rise, looks to the phonological cohort somewhat increase, and looks to the other items drop off, because they are incompatible with the firsts sounds of the noun. Eventually, as a word like “lettuce” disambiguates from cohort “letters,” looks to everything but the target drop away. This describes the proportion of looks at about 500 ms after noun onset, and reflects phonological cohort effects widely found in the word-recognition literature.

By comparison, participants have already begun to use information present in the specific measure phrase before they hear the noun at all. Looks to the target
Figure E.3  Specific English Measure Phrase Eye-Movement Proportions: Proportion of all looks to the four picture types over time (ms) in the specific condition, aligned to noun onset.

begin to rise almost immediately after noun onset, which is faster than it would take to plan a saccade in response to information in the noun; these looks are rising as a response to the measure phrase. By 300 milliseconds post onset, the target group is receiving the majority of looks. Interestingly, at no point does the phonological cohort show a strong advantage over the fourth, unrelated item. Even though the initial phonology of the target noun is consistent with that cohort, it is not a competitor for selection in the presence of a classifier that restricts it.

Since the phonological cohort is ruled out as a competitor by the measure phrase, the pragmatic point of disambiguation is earlier than the phonological POD. This manifests as quicker target selection in the specific condition, as shown in Figure E.4.

This study provides evidence that, in English, measure phrases that semantically partition collectives or substances for reference have an influence on the earliest moments of sentence comprehension. These effects mediate phonological cohort
competition when the measure phrase is inconsistent with or atypical of the cohort\(^1\).

But there is reason for caution when interpreting these results: in English, measure words often resemble nouns. In the example we have been discussing, it might seem odd to refer to a head of lettuce as simply “a head,” because there is a more dominant sense of the noun “head.” But for many other measure words in our study, including “carton,” “bottle,” and “cup,” the target picture could, if in isolation, be named felicitously with this measure term alone. Because of the temporal nature of speech, the syntax of a measure phrase and that of a mere container noun are temporarily ambiguous in English. While prosodic factors and experience with earlier trials make it unlikely that subjects believe “cup” is the end of an intonational phrase, we cannot discount the role this ambiguity might play in the strength of these effects.

\(^1\)It is also notable that in Figure E.3 the classifier competitors are more attractive initially than the target pictures. In order to avoid giving target selection a confounding advantage, when in doubt, we erred on the side of assigning the better classifier-fitting object as the competitor and not the target. Therefore the initial advantage of the classifier competitor relative to the target in the specific condition is likely to be a product of our items.

**Figure E.4** English Measure Phrase Target Looks: Proportion of all looks to the target groups over time (ms) in the two conditions, aligned to noun onset.
Chinese shares this problem of lexical ambiguity: many (especially mass) classifiers are identical in form, and clearly related, to nouns. However, Chinese syntax disambiguates a classifier phrase from a noun phrase at an earlier point in time: only a classifier, and not a noun, can directly follow a numeral. This property will allow us to test more clearly not only the influence of classifiers on language processing, but the structural level at which this takes place. Chinese massifiers may behave very much like English measure phrases, due to reference to external configuration properties; or they may behave more like grammatical gender determiners, because they are function words required by the grammar for reference to a nominal concept.

E.4 Experiment Two: Chinese Massifiers

The second study replicates the methods of the first study with Chinese mass classifiers, or massifiers. Many of these massifiers are analogous to English measure phrases. One crucial difference is that, unlike the nouns that can share their form, massifiers can occur immediately following a numeral or demonstrative term. Thus the presence of a numeral grammatically dictates that the following word cannot be a noun. Whether this cue is used by comprehenders during processing has never been investigated, so Experiment Two allows us to probe measure phrase comprehension at the level of syntactic processing.

As mentioned, some previous research has examined how Chinese mass classifiers are paired with potential referents (e.g. Li et al. (2008)). Those data suggest that comprehenders make use of ontological information in offline judgments about massifier-noun pairs. This experiment investigates whether grammatical massifiers impact comprehension online as an utterance unfolds.
E.4.1 Methods

Participants

All 18 subjects in this study were native speakers of Mandarin Chinese from the University of Rochester community who had normal or corrected-to-normal vision and no reported hearing loss. All were also minimally proficient in English, and many spoke a second Chinese dialect. Some of these participants also participated in Experiment Three with count classifiers. Subjects received payment for participation.

Materials

Items in Experiment Two closely resembled the stimuli from Experiment One. An imageable mass or massified target was chosen (e.g. *baihehua*, “lily(s)”) along with a classifier-consistent competitor (e.g. *meiguī*, “rose(s)”). A phonological competitor that shared the first syllable as the target was then selected (e.g. *baicai*, “cabbage”). Whenever possible, the cohort also shared the initial tone, but this was not the case for every item. A full list of items can be found in Appendix APPENDIX. Crucially, the phonological competitor did not typically pair with the target massifier.

Objects were displayed in much the same was as in Experiment One, as shown in Figure E.5.

One important change from Experiment One is the way the fourth picture was chosen. Instead of a phonological cohort of the classifier competitor, a cohort of the massifier term itself was chosen. For example, the classifier SHU (“bouquet”) is appropriate for the pictures we selected of roses and lilies, and so the fourth picture group is *shu*, “tree(s),” which is a phonological match for the measure word. The purpose of this was to explicitly test whether the ambiguity between noun and classifier form can account for any rapid looks to the target, even when a noun should be ruled out by the syntactic context. If so, we might also expect increased looks
Figure E.5 Chinese Massifiers Visual World Computer Display: Screen shot of Chinese mass classifier SHU referential selection trial, which shows four groups of pictures: target baihehua, phonological cohort baicai, classifier competitor meigui, and classifier-cohort shu, and a central fixation cross.

to this fourth item, the classifier-cohort, in the specific condition. Hearing classifier SHU might drive attention toward the homophonous noun shu, even though a noun is not licensed by the syntax at that point in the utterance. If we found this to be the case, this would reveal that hearing a form phonologically consistent with a picture is enough for participants to look at that item; in other words, this would lend support to a superficial explanation of classifier interpretation. If, however, subjects do not look at the picture of the trees when hearing the classifier SHU, this could provide evidence that classifiers are being treated in a syntactically appropriate manner from the earliest moments in processing.

In this study we made half of the target items true mass substances and the other half massified groups of discrete objects. Because Chinese does not syntactically
distinguish mass and count nouns, we did this based on perceptual intuitions as well as by analogy to items in the preceding English study.

A naive female native Mandarin speaker was chosen to act as the voice for our experimental instructions. We recorded a general instruction, such as xuan yi XIE baihehua, which is translated roughly as “choose some lilies” or “choose a few lilies,” and contrasted that with a more specific instruction, xuan yi SHU baihehua, literally “choose one BOUQUET lily(s),” or “choose a bouquet of lilies.”

Fourteen fillers were included that appeared identical to critical items, except that the target was one of the other three item types in the display. Half the fillers used the general classifier XIE in the instruction, and the other half used a specific massifier appropriate for the referent. As in Experiment One, this was done to prevent subjects from learning that the target was a picture that had both a classifier and a phonological competitor co-present in the display.

**Procedure**

The procedure was identical to that in Experiment One.

**E.4.2 Results**

The data patterns in this study were similar to the English massifier data. Looks to target items do not increase until 550 milliseconds after noun onset in the general condition (shown in Figure E.6), whereas they rise prominently within 150 milliseconds of noun onset in the specific condition (shown in Figure E.7). Phonological cohort competition, while never especially prominent, is nearly twice as large in the general condition. The classifier-cohort items receive very few looks overall.

**E.4.3 Discussion**

Effects of the massifier on eye-movements occurred very early, before information contained in the noun itself could direct a saccade. This contrast between the gen-
APPENDIX E. INCREMENTAL INTERPRETATION OF CHINESE CLASSIFIERS AND MASSIFIERS

Figure E.6 General Chinese Massifiers Eye-Movement Proportions: Proportion of looks to the four picture types over time (ms) in the general condition, aligned to noun onset.

The attention directed to phonological competitors was diminished in the specific condition, where the massifier was inconsistent with the cohort. This suggests that semantic restrictions in the massifier are excluding the cohorts as potential referents, even when phonological information in the noun phrase is consistent with those pictures. Most noticeably, we find a target advantage in the specific condition, where the information in the massifier speeds referential selection.

As previously mentioned, these effects alone do not speak to the level of processing that is taking place. To better understand the level of processing at work, we included the classifier-cohort objects, like *shu*, “tree,” which share phonology (although not necessarily orthography) with the specific classifier used, which is...
SHU ("bouquet") in this example. If looks had increased to the classifier-cohort in the specific condition, this could have indicated that comprehenders are using massifier information at a superficial level, where the phonological consistency of a small window of input directs attention to matching pictures. If this had occurred, part of any effects we found might be due to a boost when the classifier shares its form with a related noun that could also describe the target picture. An example of this in English would be that “bottle” could felicitously refer to a picture of wine, if the wine is depicted as being in a bottle.

However, looks to the classifier-cohort were always minimal. They were, in fact, lower in the specific case, where the matching phonology was present, than in the general condition, where XIE was always used as the classifier. This is the pattern of results we expect if the presence of yi, the numeral, provides a clear signal to comprehenders that the next thing they hear cannot be a noun, but is likely to be a classifier. In other words, these data indicate that subjects interpreted the classifier terms in a structurally relevant way, and not as superficial phonotactic signals, nor
as potential nouns.

These data show that massifier affordances rapidly influence a comprehender’s set of referential alternatives with mass referents. These effects are operating at a syntactic level. The results of Experiment Two highlight the similarity of Chinese massifiers to English measure phrases: both seem to partition a set portion of a substance, or name a collective. The comprehender is able to use the information in the measure phrase very quickly to select a target before its phonological point of disambiguation. Nouns that sound the same but are partitioned differently do not compete during referential selection. However, we know intuitively that this type of semantic circumscription is more necessary when referring to mass or massed referents than count referents. Thus count classifiers might not provide information that is important during sentence processing. On the other hand, count classifiers are required by the grammar of Chinese and often evoke conceptual features of the referent. Experiment Three directly assesses the influence of count classifiers on Chinese sentence comprehension.

E.5 Experiment Three: Chinese Count Classifiers

Experiment Three directly investigates the influence count classifiers have on Chinese sentence processing when potential referents appear as individuals, not groups or substances. Huettig et al. (2008) and Tsang and Chambers (2008) both use eye-tracking to study Chinese count classifiers. Yet both groups are concerned with absolute looks to one particular item in the display. Neither group explicitly looks at how the level of classifier informativity can impact the dynamics of competition between objects during referential selection.
E.5.1 Methods

Participants

Eighteen native Mandarin speakers from the University of Rochester community participated in this study for pay. All had normal or corrected-to-normal vision and no reported hearing problems. All were at least minimally proficient in English, and many spoke other dialects of Chinese in addition to Mandarin.

Materials

Stimuli in this study closely resembled those of Experiment Three with one major difference: all items displayed were conceptual count objects, not mass substances, and none of the objects appeared in cohesive groups (i.e. no “bouquets” of flowers, “packs” of cigarettes, et cetera).

Target items, such as *men* (“door”) were paired with words that take the same classifier, like *chuanghu*, (“window”). Phonological cohorts (*menpiao*, “ticket”) shared an onset syllable (and, as often as possible, tone) with the target but did not pair with the target classifier. The fourth item again was a phonological cohort of the specific classifier itself. In this item, *men* and *chuanghu* both pair with the classifier SHAN, which connotes flat vertical things, so the classifier-cohort was *shanzi*, “fan.” To preserve felicity and to maintain the same level of visual complexity as in Experiment Three, the individual objects were depicted in groups of three identical tokens, as in the first two experiments. This example is shown in Figure E.8.

Again we compared a specific instruction, such as *xuan yi SHAN men*, “Choose one CL-flat door,” with a less informative instruction. Chinese has a general classifier for individuals, GE, which we used in lieu of the restrictive classifier in our general instructions, as in *xuan yi GE men*, “Choose one INDIVIDUAL door.” GE has been argued to be a true default classifier (Myers, 2000), although acceptance of GE with a particular noun may vary, especially in formal speech or writing. More importantly, GE seems the best uninformative pragmatic equivalent for the selection of individuals.
APPENDIX E. INCREMENTAL INTERPRETATION OF CHINESE
CLASSIFIERS AND MASSIFIERS

Figure E.8  Count Classifier Visual World Computer Display: Screen shot of Chinese count classifier SHAN referential selection trial, which shows four groups of pictures: target men, phonological cohort menpiao, classifier competitor chuanghu, classifier-cohort shanzi, and a central fixation cross.

A female native speaker of standard-sounding Mandarin recorded all instructions. Fourteen fillers, which were randomly mixed with critical trials, were of the same type as in the previous study.

Procedure

The procedure was identical to that of Experiment Two.

E.5.2  Results

As in Experiment Two, we are concerned with looks to each item type as information from the speech stream is processed by the comprehender. Figures E.9 and E.10 show the proportion of looks to each item type as the utterance unfolds.
In the general condition, when the default classifier GE was a good match for all the items shown, we see looks to all four object groups attract approximately equivalent looks until around 200 ms after the noun onset, at which point the classifier competitor and classifier-cohort looks begin to drop. The phonological cohort looks remain steady and target looks rise until around 400 ms post noun onset, at which point the phonological competitor looks begin to recede. Target items attract half of all looks by about 550 milliseconds after the noun begins.

The specific classifier condition looks different. Initially looks to all four object groups are equivalent, but around 100 milliseconds after noun onset, looks to the phonological competitors begin to fall. The classifier-consistent competitor looks peak near 30 percent at approximately 350 milliseconds after noun onset. By 250 milliseconds after noun onset, the target group is attracting nearly one third of looks, and by 450 ms post onset that becomes half of all looks.
Figure E.10  Specific Chinese Classifiers Eye-Movement Proportions: Proportion of looks to the four picture types over time (ms) in the general condition, aligned to noun onset.

E.5.3  Discussion

The basic pattern of these results resembles the massifier data: phonological cohort and classifier competitor trade off as the dominant challenger for attention depending on the level of classifier restrictiveness. The target receives the majority of looks about 100 milliseconds quicker in the specific classifier condition. This advantage demonstrates that comprehenders make swift use of nominal classifiers during reference resolution. As with massifiers, the classifier-cohort items earn very little attention from subjects, suggesting that facilitation effects from the specific classifiers cannot be attributed simply to phonetic attention-driving.

Notably, competition from the phonological cohort practically disappears in the specific condition. This is consistent with Dahan:2000hl research on grammatically gendered determiners. However, this pattern is not as strong with individuated as with massified referents. Classifier competitors looks drop away later and to a lesser degree with count classifiers than with massifiers. In other words, the degree to
which the classifier competitor emerges as a challenger in the specific condition is not as strong. The pattern, however, is still important and indicates that, first of all, classifier specificity matters, and secondly, count classifiers are playing a role in referential selection by changing the set of competitors that are entertained during comprehension, as well as by somewhat facilitating target selection.

E.6 Comparing Across Language and Ontology

Because the procedures used in the three studies were nearly identical, differences in the data provide insight into classification across languages and ontologies. What we see in broad strokes is that classifiers across languages and types provide contextual information that is used in the earliest moments of sentence comprehension, although the impact and nature of these effects varies somewhat by language and ontology.

With English measure phrases, we see a clear early influence of measure words on anticipatory looks toward classifier-consistent objects. In the specific condition, just before the onset of the noun, only about twenty percent of looks are to objects inconsistent with the measure phrase. Notably, the target’s phonological cohort is only a competitor when the measure word is generic enough to allow it to be a plausible target. This is unsurprising given our intuitions that English measure words are contentful. This effect is likely strengthened by the ambiguity between measure words and related nouns denoting containers in English.

Chinese massifiers play a similar role in dividing up the world for counting, but they appear in grammatical structures that identify them unambiguously as classifiers, not nouns. While some massifiers share morphology or form with nouns, these nouns cannot directly follow a numeral term, but would need their own classifier. The absence of looks to the classifier-cohort pictures indicates that comprehenders are aware of the syntactic category of the massifier and are using this information during comprehension. Therefore, early looks to the target and classifier competitors in the specific condition are not attributable to the class ambiguity of some
massifiers, or to superficial phonological attention-driving.

Although Chinese massifiers have meanings that seem to map well to English measure phrases, their impact on comprehenders’ visual attention is not quite as strong. In the specific condition, where the massifier is informative enough to halve the number of potential targets, looks to all four object types are approximately the same until shortly after noun onset. This represents a delay compared with English measure phrases, although these effects still occur before information in the noun could be used to plan and launch a saccadic eye-movement. And as in Experiment One, the strength of the phonological cohort competition varies based on massifier specificity.

Massifier information in Chinese constrains comprehension quite rapidly, both by facilitating target selection as shown in Figure E.11, but also by influencing which other items in the display compete with the target. The delay of these effects relative to English, however, raises the possibility that massifiers in Chinese direct comprehenders’ attention to consistent objects somewhat less powerfully.

Count classifiers show even more subtle effects in the same direction. The target has very little advantage in the presence of a more informative classifier, as shown in Figure E.12. However, the type of object that competes the most strongly for selection varies with classifier specificity. Count classifiers, then, seem to offer contextual information that constrains comprehenders’ interpretations. However, the advantage in terms of target selection is minimal. Although count classifier/noun relations are more limited by Chinese grammar, and might therefore provide clearer information about what nouns can come next in the utterance, our data show that they offer less facilitation for target selection than massifiers or English measure phrases do.

We must also examine alternative explanations for the differences we see in our data than an ontological distinction encoded by the classifier. One possibility is that the lexical co-occurrence of our Chinese mass nouns with our massifiers was higher than the co-occurrence of our count classifier/noun pairs. If the probability of a certain noun given a classifier was higher overall for the mass nouns in our study,
this could result in the faster and larger influence on comprehension of massifiers relative to classifiers.

We preliminarily investigated this by searching for classifier and classifier/noun bigram frequencies on Google’s Chinese language pages. We then divided the classifier frequency by the bigram frequency to quantify the proportion of times the noun appears when a particular classifier is given. A t-test between the log values of these proportions revealed no significant differences between the predictiveness of our massifiers and count classifiers. If anything, the count classifiers are more predictive of our count referents than our massifiers are of our mass referents by this metric ($t = -1.22, p = .232$), but this is not statistically significant\(^2\).  

There are additional reasons besides these co-occurrence data to believe that count classifiers maybe be generally more predictive than massifiers. Massifiers can be paired with many nouns and still maintain grammaticality and interpretabil-

\(^2\)Despite this preliminary finding, we do not want to completely rule out the possibility that, given a large spoken corpus of Chinese and further analyses, classifier/noun frequency information can be used during online processing. We hope to address this in more detail in future work.
Meanings of even odd massifier/noun pairs can be coerced, and new containers or quantities can be created and turned into massifier terms. Count classifier/noun pairings, however, are more grammatically rigid, and the use of a count classifier with a mismatched noun is typically unacceptable to native speakers. Thus massifiers are an open and flexible class, and are therefore plausibly less predictive than count classifiers. Moreover, according to Erbaugh (1986), only around three percent of nouns referring to physically co-present entities appear with a non-generic classifier in conversation. So it is possible that, while specific count classifiers are more restrictive than massifiers, comprehenders are less accustomed to having this information available during referential selection.

The fact that the massifiers prompt a stronger influence on comprehension in the visual world, despite being less restrictive, could be explained by another salient difference between measure words and massifiers on the one hand and count classifiers on the other: mass classifiers tend to align more with external visual properties than count classifiers do. That is, were a substance like milk to be arranged in visually distinct forms, it would likely be best denoted by different massifiers, as in “a glass of milk,” “a carton of milk,” or “a puddle of milk.” A count object like
snake, however, can be shown in several different poses and still be paired with the same classifier in Chinese. In other words, massifiers may be more informative with regard to the visual world paradigm we used in these studies. Nevertheless, the ontological differences we found support the observation that massifiers refer to external containment or configuration of an object, which may be transient properties, while classifiers refer to physical or functional properties internal to the object’s structure, which are permanent (Tai, 1994).

E.7 Conclusions

In these studies, we find rapid effects of measure words, massifiers, and count classifiers on sentence processing in English and in Chinese. We have demonstrated that these words are being processed in a syntactically relevant way, and not merely exerting a superficial influence on comprehension. Our data also reveal slight differences between mass and count in Chinese, where this distinction is not syntactically realized. However, these differences may be driven by the visual nature of the task, and may reflect a general pattern in the language where massifiers encode more visually salient information than classifiers.

The gradient facilitation we find from English measure phrases to Chinese massifiers, and finally to Chinese count classifiers is in accord with cognitive intuitions as well as linguistic observations that these words may contribute different types and degrees of information, although further research is necessary to fully explicate the mechanisms of mass and count reference across languages.

---

3Unless, of course, the snake is crushed into a substance or is a member of a collective group