Reference Resolution and Discourse Salience

by

Mikhail Masharov

A Dissertation
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
The College
Arts and Sciences

University of Rochester
Rochester, New York
2008
Curriculum Vitae

The author was born in Novosibirsk, Russia. He has attended institutes of higher education in Russia, New Zealand and the United States. The author attended the Department of Applied Mathematics at Novosibirsk State University, Russia from 1992 to 1994. He then attended University of Auckland, New Zealand from 1995 to 1996 and graduated with a degree of Bachelor of Science in Computer Sciences with Honours in 1996. The author came to the University of Rochester in the Fall of 1999 and began graduate studies in Brain and Cognitive Sciences. He pursued his research in language processing under the direction of Professor Michael Tanenhaus, and received the Master of Arts degree from the University of Rochester in 2004.
Acknowledgements

First and foremost I would like to thank my advisor, Mike Tanenhaus. His never-ending support and patience, extensive knowledge, enthusiasm for everything he does, and encouragement have been instrumental over the years in pointing me in the right direction. I am also indebted to Elsi Kaiser and to the members of my committee Greg Carlson and Florian Jaeger for their constructive comments, suggestions and criticism of the various stages of the work presented in this thesis.

I am also sincerely proud to have been a part of BCS and believe that nowhere else would I have found such an amazing collection of open-minded and brilliant people! I am grateful to everyone in BCS, CLS & CVS for making my experience in Rochester a stimulating one!

Some people I have met here deserve a special mention. Mrim Boutla – for the endless sleepless nights in the office and the discussions of the true meaning of BCS (Baking & Cooking School). Also, for seriously impeding my progress towards a dissertation by presenting me with a book of a thousand recipes. Joszef Fiser – for a unique and frank perspective on the life of research. Helena Kopecky, Laura Silverman, Carla Hudson – for being good friends. The Sarahs – Sarah Creel & Sarah Brown-Schmidt – for being great officemates! Jim Magnuson, Michael Spivey, Jennifer Arnold – their enthusiasm is contagious!
Finally, I would like to thank people without whom this dissertation might never have been accomplished. Dana Subik – for all the help in the lab and with running experiments. Without his help this work would not be possible. My Russian friends – Filipp & Yana Ignatovich, Julia Kazdobina – for their companionship, good humor and late night discussions! Спасибо, что вы есть! A very special thank you to Svitlana Gubriy for always being there for me. Your support, patience and encouragement are priceless to me!

Last, but not least – a thank you to Lila Gleitman and Tom Wasow – for suggesting Mike as an advisor when I was looking for one.

This work was supported by the National Institute of Health Grant HD-22067 to M.K. Tanenhaus.
Abstract

This dissertation investigates predictions from the Givenness Hierarchy and similar theories that associate different levels of discourse entities’ salience with different anaphoric forms. The experimental methodology utilizes the visual world paradigm to assess referent preferences of the three chosen anaphoric forms (indefinite, ‘a(n) N’; definite, ‘the N’; and demonstrative, ‘that N’) under conditions of developing discourse. The discourse entities’ salience is manipulated through the introduction of varying degrees of linguistic material between the antecedent and the subsequent reference. Participants followed auditory instructions to click with a computer mouse on various objects on the screen, allowing us to analyze both the online reference resolution process (eye gaze fixation proportions) as well as their final referent choices. We show that only discourse-relevant information affects discourse entities’ salience, that different referential forms react differently to such salience changes, and that the initial level of salience affects this sensitivity to change.

Overall, our findings are problematic for approaches that use a single dimension, such as salience, to associate different referring expressions with their preferred referents. The framework that best accommodates our results is the form-specific multiple-constraints approach advocated by Kaiser and colleagues, which proposes that while an entity’s salience is important for determining the appropriate reference, individual anaphoric forms carry their own specific constraints on the conditions of their felicitous use.
Experiments 1 and 2 establish the methodology and examine the change in the preferred referent for different referential forms under changing discourse conditions. We also examine how the initial salience of a discourse entity affects reference resolution by having the discourse introduce a potential referent as either the Theme (Experiment 1) or the Goal (Experiment 2). The results demonstrate that different anaphoric forms are affected differently by the dynamically changing discourse conditions. The initial preference for all anaphors was for the salient discourse mentioned entity, with increasing intervening material facilitating acceptance of the low-salience unmentioned entity as the intended referent. This effect was the strongest for indefinites and the weakest for demonstratives.

Experiment 3 investigates possible reasons for the preference of indefinites for the mentioned referents. The experiment evaluated the hypothesis that an indefinite prefers to refer to one of several low-salience potential referents by creating a situation where the set of possible referents includes two identical low-salience items. The results of this experiment largely replicate the results from Experiments 1 and 2, which suggests that the indefinite form is interpreted in line with an ‘any’ interpretation rather than in line with the Givenness Hierarchy predictions.

In Experiment 4 we investigate the same questions with more structured, schematic scenes, a narrative discourse, and a scene verification task. We use dynamic scene changes during the narrative description to manipulate the
alternatives that participants had to consider in making their response. The results confirm earlier findings – the indefinite form resolved equally well to mentioned and non-mentioned referents, while the demonstrative reference did not accept a non-mentioned entity as a possible referent.

We also investigate the possibility of evaluating the relative salience of discourse entities directly by using a change blindness paradigm. This manipulation suggests the feasibility of monitoring discourse representations through non-linguistic means.
# Table of Contents

## Chapter 1  Introduction  
1  
- Accessibility Theory  
- Salience Hierarchies  
- Multiple-Constraints Salience Model  
- Defining Salience  
- Information Load Theory  
- Centering Theory  
- Other Approaches  
- Overview of Thesis  

## Chapter 2  Experiments on Dynamic Changes in Discourse Salience  
26  
- Experiment 1 – Theme  
- Discussion  
- Experiment 2 – Goal  
- Discussion  

## Chapter 3  Processing of Indefinite Referential Form  
71  
- Experiment 3 – 3 lemons  
- Discussion  

## Chapter 4  Reference Processing in a Narrative Discourse  
86  
- Experiment 4  
- Discussion  

- Non-linguistic Assessment of Discourse Salience  

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discussion</td>
<td>123</td>
</tr>
<tr>
<td>Chapter 5</td>
<td>125</td>
</tr>
<tr>
<td>Summary and Conclusions</td>
<td>125</td>
</tr>
<tr>
<td>References</td>
<td>130</td>
</tr>
</tbody>
</table>
List of Tables

Table 2.1: Givenness Hierarchy and Associated Anaphoric Forms in English 27
Table 2.2: Sample Discourse, Experiment 1 31
Table 2.3: Sample Discourse, Experiment 2 58
Table 4.1: Sample Discourse, Experiment 4 92
List of Figures

Figure 2.1: Sample Display, Experiments 1 & 2 30
Figure 2.2: Referent Choices, Experiment 1 44
Figure 2.3: Fixation Proportions during Discourse Setup, Experiments 1 & 2 47
Figure 2.4: Fixation Proportions on the Mentioned Referent 48
Figure 2.5: Fixation Proportions on the Non-Mentioned Referent 49
Figure 2.6: Referent Choices, Experiment 2 61
Figure 2.7: Fixation Proportions on the Mentioned Referent 62
Figure 2.8: Fixation Proportions on the Non-Mentioned Referent 63

Figure 3.1: Sample Display, Experiment 3 73
Figure 3.2: Referent Choices, Experiment 3 75
Figure 3.3: Fixation Proportions on the Mentioned Referent 76
Figure 3.4: Fixation Proportions on the Non-Mentioned Referents 77

Figure 4.1: Sample Display, Experiment 4 89
Figure 4.2: Percent of ‘Match’ Responses to Mentioned Changes 102
Figure 4.3: Percent of ‘Match’ Responses to Non-Mentioned Changes 102
Figure 4.4: RTs of ‘Match’ Responses to Mentioned Changes 103
Figure 4.5: RTs of ‘Match’ Responses to Non-Mentioned Changes 103
Figure 4.6: RTs of ‘No Match’ Responses to Non-Mentioned Changes 104
Figure 4.7: Fixation Proportions during Mentioned Change Trials 110
Figure 4.8: Fixation Proportions during Non-Mentioned Change Trials 111
Figure 4.9: Sample Display – Color Change, Experiment 4 116
Figure 4.10: Fixation Proportions during Color Change Trials 119
Figure 4.11: Mean Fixation Proportions 120
Figure 4.12: Percent of ‘Match’ Responses on Color Change Trials 121
Figure 4.13: RTs of ‘Match’ Responses on Color Change Trials 121
Chapter 1 – Introduction

Whenever we engage in conversation, one of the main things that speakers and listeners have to do is to structure the unfolding discourse in a meaningful way. And whenever people talk, they refer to events and entities in the world. Interlocutors have to understand what is being talked about and integrate any new information with their representation or the mental model of the discourse. This process of reference resolution is one of the central aspects of language use. Referential expressions play an important role in shaping natural languages into a powerful communication tool that they are – they maintain recursive properties of languages, help avoid redundancy and shorten/simplify utterances. Thus, in order to understand the nature and mechanisms of language processing, it is essential to understand how reference is resolved. Important questions to examine include how different referential forms are linked to discourse entities and what mechanisms are involved in tracking and modifying the representations of these entities over time.

Throughout this dissertation the term “discourse” will be used to refer to a coherent set of utterances, where coherence is meant to indicate some sort of common purpose or information of this set of sentences. The term “anaphor” will be used for referring expressions that refer to an entity that has been previously invoked in the discourse. An “antecedent” will refer to the linguistic form with which that entity was last mentioned, and “referent” will mean the actual entity or its conceptual representation that the anaphor refers to.
Usually, there is more than one anaphoric form that can be used to successfully refer to a discourse entity. For example, I can refer to a pencil as “the pencil”, “a pencil”, “that pencil”, “that one”, “that” or “it”. A central problem for linguists and psycholinguists is to understand the factors that affect the speaker’s choice of a particular referential expression in any given circumstance, and the factors that affect how the listener recovers the speakers intended referent when he or she encounters a referring expression. In most cases, anaphors appear in the same discourse as their antecedents, but not within the same sentence – the domain over which syntactic rules apply. Therefore, these discourse anaphors are not bound by syntactic rules that render some forms grammatical and others not, but instead seem to be bound by a set of constraints that make certain forms “better” or more “felicitous” than others in a given situation (Ariel, 2001; Gundel et al., 1993). A general consensus in the field is that the wide variety of possible referential forms and the appropriate situations of their usage are derived from a differential cognitive status of the referent (Ariel, 1990, 2001; Givon, 1983; Gundel et al., 1993; Prince, 1981). This view assumes that referential forms are subject to a discourse salience or accessibility ranking which determines the appropriate reference form for any given situation in a discourse.

Early work on salience and the production of referring expressions focused on perceptual salience and the importance of the information to the speaker. For example, Osgood and Bock (1977) identified three kinds of salience: naturalness – the natural order of constituents mirrors the natural order of events (Agent-action-
Patient), vividness – the inherent salience of the semantic features of the entity (e.g. the murderer vs. the man); and motivation of the speaker – salience attributed by the speaker (such as interest, concern, perspective).

While perceptual salience clearly plays some role in the choice of what to refer to and how to refer to it, it is likely to be most useful in descriptions of scenes or in interpreting sentences out of context. The situation is likely to be different in discourse – connected text or dialogue – where the choice of what gets mentioned first and how to refer to it is likely to be at least partially determined by what was mentioned in the preceding discourse, and how salient it is at the time that the speaker chooses to refer to that information.

**Accessibility Theory**

Attempts to associate a referring expression with some kind of cognitive status of its referent have a long history starting with Chafe (1976), to the Familiarity Scale by Prince (1981), to Givon (1983), to what is perhaps the most detailed early theory of discourse salience and its relevance to reference resolution – the Accessibility Theory by Ariel (1990). According to Ariel, “accessibility theory offers a procedural analysis of referring expressions, as marking varying degrees of mental accessibility” (Ariel, 2001). Here, referring expressions are viewed as instructions to the listener to retrieve a certain piece of information based on how accessible it is. The level of accessibility is presumed to be coded by the referential form itself – as an accessibility marker of sorts. Also, importantly, the theory claims that it is the
discoursal rather than the physical salience of the entities involved which determines
the degree of accessibility assigned to particular mental representations (Ariel,
2001).

Since then many functional linguistic theories assume that the form of
anaphoric expressions signals the listener how accessible or salient the antecedent is
(Ariel, 1990, 2001; Gundel, Hedberg & Zacharski, 1993). For example, pronouns
signal that the antecedent is highly accessible, whereas full noun-phrases (NPs) and
names signal that the antecedent is relatively inaccessible (Gordon, Grosz & Gilliom,
1993). In addition to marking a current degree of activation, the form of NPs also
partially determines the degree of activation subsequent to their utterance — referred
to as future accessibility by Ariel (2001). In short, the more explicit an NP is, the
greater the subsequent increase in activation of the corresponding referent. In support
of this view, Gernsbacher (1989) presents evidence that proper names reactivate an
antecedent more strongly than a pronoun. From this perspective, current activation
marking is in an inverse relation to future activation marking. A higher accessibility
marker like a personal pronoun indicates high current accessibility, but does
relatively little to increase activation.

However, the fact that reference resolution happens online and is based on a
rather limited available information suggests a mechanism that does not derive the
meaning of reference exclusively from the form used, but rather draws on a richer set
of mental representations constructed during prior discourse. At the same time all
referential expressions are interpreted with respect to some domain containing the set
of possible referents. Selective attention can either limit the domain of possible linguistic referents (Brown-Schmidt & Tanenhaus, 2008) or shift the resolution process from one domain into another. For example, a disfluent production biases the listener’s initial preferred interpretation from an entity that has been mentioned (discourse-given) to a previously unmentioned or (discourse-new) entity (Arnold et al., 2004). Different referring expressions seem to be restricted to certain types of entities. For example, the current focus of attention limits the domain of interpretation of a pronoun (e.g., “it”) to the currently attended entities. Similarly, a definite reference (e.g., the pencil) limits it to the set of all uniquely identifiable entities, and an indefinite referring expression to the set of all conceptually or type identifiable entities. Greene, McKoon and Ratcliff (1992) argue that research on pronoun resolution must consider the discourse contexts in which pronouns are likely to occur. Any theory of reference resolution has to take into account the mental model of the prior discourse and of the current context.

It is assumed that discourse processing involves construction of a mental representation of the discourse elements (Gernsbacher, 1989; Kintch, 1988; Morrow et al., 1987). Referential forms require the entity to which they refer to be activated in memory to some degree. This level of activation is often called the mental salience of an entity. Each entity is assumed to have some degree of accessibility, which is determined in part by the syntactic and semantic structures in which it is linguistically expressed. Accessibility is measured relative to the local environment, that is, relative to the other entities introduced in nearby clauses and sentences.
(Greene, McKoon, & Ratcliff, 1992). As the reader or listener moves through a discourse, the accessibility of entities changes as the local environment changes. The entity or entities that are most accessible at any point are what the discourse is about at that point, i.e. the focus or topic.

This notion of salience in language and discourse has been around for a long time, but has been notoriously difficult to define. It is sometimes used interchangeably with focus, accessibility, prominence, activation, retrievability, availability, etc. In any case, the cognitive status of a referent described in these terms is assumed to partly determine the anaphoric form used to access that referent.

The goal of much past research has been to identify specific factors that affect the use of referential forms. Adopting a partial listing from Arnold (1998), these factors include recency of mention, topicalization (influenced by subjecthood), first-mention, repeated mention, word order, prior pronominalization, focus (syntactic constructions or intonation), semantic properties (thematic role, event structure, etc.). Studies investigating language production have found that concrete nouns and animate entities (due to their inherent conceptual salience) have a greater tendency to appear early in sentences as the subject (Bock & Warren, 1985; McDonald, Bock, & Kelly, 1993). Pearson, Poesio and Stevenson (2001) also find that animacy, thematic role (Theme and Goal) and recency affect what is mentioned first. Some theories assume that the choice of a form is determined by the strongest applicable salience factor (McKoon & Ratcliff, 1992; Gernsbacher, 1989), while other theories adopt a perspective of multiple factors combining with each other in
influencing the referent’s salience level (MacDonald & MacWhinney, 1990; Hudson-D’Zmura & Tanenhaus, 1998). Recency of mention and topicalization are considered the strongest factors in focusing the reader’s or listener’s attention on the discourse topic (Arnold, 2001). References to recently mentioned entities are usually processed more easily and faster than references to more distant antecedents (Arnold, 2000a). Similarly, topicalization of the antecedent serves to put it into mental focus and so influences the ease of reference resolution (Garrod, Freudenthal & Boyle, 1994). Other factors have less weight, but can bias the interpretation in specific circumstances.

Salience Hierarchies

Having identified individual factors affecting the use of anaphors, several theories have attempted to provide a comprehensive functional account of the relationship between a referring expression and the cognitive status of the discourse entity that it refers to. One of the main theoretical constructs is that of a salience hierarchy (Ariel, 1990, 2001; Gundel et al., 1993) which ranks referential forms in terms of how restricted they are in their usage. Such a ranking assumes that referential forms vary according to the salience of the referent. Generally, with pronouns and null forms referring to the most accessible entity and with definite and indefinite expressions referring to increasingly less accessible entities. Overall, there is a general consensus that the more reduced an anaphoric expression is, the more salient the entity it refers to has to be (Givon, 1983). Critical evidence in favor of
such structures comes primarily from intuitions about felicity and from corpus
analyses. Perhaps the most explicit theory is the Givenness Hierarchy by Gundel et
al. (1993), which differentiates six levels of salience corresponding to different
hypothesized cognitive states. The six levels (top row) and the referential forms
(bottom row) that minimally require each of these levels are presented below:

<table>
<thead>
<tr>
<th>in focus</th>
<th>activated</th>
<th>familiar</th>
<th>uniquely identifiable</th>
<th>referential</th>
<th>type identifiable</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>it</em></td>
<td><em>that, this</em></td>
<td><em>that</em> N</td>
<td><em>the</em> N</td>
<td><em>indefinite</em></td>
<td><em>a</em> N</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>this N</th>
</tr>
</thead>
</table>

One general problem that arises in linguistic work on the relationship
between referring expressions and the proposed salience or accessibility is that
salience of an entity introduced by the linguistic antecedent is used to explain the
felicity of different anaphoric forms, while at the same time the preferred anaphoric
form is used as a diagnostic for salience. For example, consider the discourse: “Lisa
picked up the bowl on the table. It had some peaches in it.” In this example, ‘it’
refers to the most salient discourse entity following the first sentence – the bowl.
However, the reason we know that ‘the bowl’ is the most salient entity is because we
can reference it with a pronoun ‘it’. Without an independent way to assess salience,
there is a clear danger of circularity that can only be avoided through an independent
non-linguistically motivated way of measuring or assessing discourse salience.

Another problem with the salience rankings approach is the lack of a clear
linking hypothesis for the correlation between a referential form and the degree of its
referent’s salience – why with an increase of the referent’s salience, the amount of descriptive information contained in the anaphor decreases. Given that these models do not provide a functional definition of salience or how it can be assessed, this informativity question prevents a comprehensive discussion of the mechanisms underlying reference resolution.

Salience rankings such as the ones discussed above generally represent attempts to structure anaphoric forms with respect to their definiteness using a single ranking factor (salience) rather than on the basis of form related properties. There are several major property distinctions or licensing conditions that have been explored with the aim to differentiate the definiteness of referential forms. Uniqueness (Russell, 1905) and familiarity (Heim, 1982) approaches to definiteness are two of the main competing theories with uniqueness being a semantic property, while familiarity is discourse-pragmatic in nature (Abbott, 2003). The analysis of definites by Russell treats uniqueness as a quantificationally expressed property, rather than one that is properly referential; indefinites are likewise quantificational but lack uniqueness. The familiarity condition, on the other hand, has been proposed by Heim (1982) as non-quantificational with the distinction between definite and indefinite descriptions being attributed to a variable on a Novelty-Familiarity continuum. Her File Change Semantics system proposed that indefinites introduce new entities, while the definites refer to discourse-old information. Both approaches, however, are unable to account for some specific examples of anaphor use, mostly involving speaker’s intentionality or specificity of the referent. A more recent alternative might
be to consider reference resolution within a wider context of both discourse and the circumstances of use as proposed by Dekker (1998). This approach suggests that contextually available information (e.g. the current task or situation) should be able to license a specific referential form along with its preferred interpretation. This would constitute a step back from the salience rankings based on a single factor and towards a more form-specific account of reference resolution.

Although there is an extensive psycholinguistics literature that examines the processing of definite pronominal reference and definite NP anaphor (Garnham, 2001, for an overview), relatively little experimental work has been aimed at validating the claims of salience ranking models and specifically the Givenness Hierarchy. One example is a study by Warren and Gibson (2002), which finds correlations between anaphoric forms and discourse status by demonstrating a differential processing cost based on the anaphoric form. They assume that reading times reflect the amount of processing resources that the listener uses to process different referring expression. The fact that changing a referring form can spur differential processing costs, and that these costs pattern as predicted by an independently established accessibility hierarchy (Gundel et al., 1993) demonstrates that the referring form is an important indicator of discourse status to the referential processing system. Warren and Gibson also extended this finding across multiple types of NPs – reading times on the main verb were fastest following a pronoun condition, slower following names, slower still for definite descriptions, and slowest
for indefinite descriptions, all conforming to the predictions of the Givenness Hierarchy.

A major assumption underlying theories of accessibility hierarchies or rankings is that salience is the main unifying dimension that distinguishes among different referential forms. However, it is unclear whether a uniform dimension such as salience or activation is sufficient to account for the differences among referring expressions. Grammatical features, such as number or gender, can influence reference resolution above and beyond discourse salience (Arnold et al., 2000b; Sanford & Garrod, 1989). Many researchers regard salience as a ‘compound’ notion resulting from the interaction of multiple factors or constraints (Ariel (1990); Arnold (1998); MacDonald & MacWhinney, 1990; Hudson-D'Zmura & Tanenhaus, 1998). Various psycholinguistic experiments (Arnold et al., 2000b; Gordon et al., 1993; Brown-Schmidt et al., 2005) also suggest that anaphor resolution can be influenced by multiple constraints.

Multiple-Constraints Salience Model

To a first approximation, there are two main approaches to referent salience:

(i) the single-factor approach, which assumes that a uniform salience level determines the form of the referring expression (Ariel, 1990; Gundel et al., 1993; Givon, 1983; Prince, 1981; etc.), and
(ii) the multiple-factor approach, which assumes that, in addition to the referent’s salience, multiple form-specific weighted constraints contribute to the choice of the referring expression used to address the
referent (for a comparison of the two approaches see Kaiser & Trueswell, 2008). The main question is whether it is plausible to rank all referential expressions of a language along a unified salience continuum.

Notably, Kaiser investigated whether different referential expressions are sensitive to different kinds of factors that go beyond simple salience, and whether this sensitivity varies depending on the referential form used. She conducted a number of cross-linguistic studies in Finnish, Estonian and Dutch to demonstrate that the predictions of the Givenness Hierarchy do not extend well to other languages (Kaiser, 2003; Kaiser & Trueswell, 2008). Using corpus analyses, questionnaires and on-line processing studies, Kaiser and colleagues demonstrate that the interpretation of certain Finnish pronouns and demonstratives cannot be fully described by a unidimensional account of salience. Instead, the results indicate that in contexts where the referents are full NPs, some anaphors are sensitive to information at the syntactic level, while others to information at the discourse level. This suggests that contrary to what is commonly assumed, referential forms cannot be mapped onto a unified salience continuum, because they exhibit varying degrees of sensitivity to, for example, syntactic role and word order.

Brown-Schmidt et al. (2005) examined the interpretation of the English pronouns *it* and *that*. They contrasted reference to a composite object (e.g. a cup on a saucer), obtained during experimental manipulations, which did not have a linguistically introduced antecedent, with references to items that were previously mentioned in the discourse. The experiments demonstrated a preference of the
pronoun *it* for a salient discourse-mentioned object. However, *that* consistently preferred composite entities, and the interpretation of both pronouns was also modulated by non-linguistic factors, demonstrating that a unidimensional account of salience cannot predict the usage of all referential forms.

Results like these, which demonstrate that different referential forms are sensitive to different levels of representations and different kinds of information, are problematic for uniform salience-based accounts. Kaiser and Trueswell (2008) strongly argue in favor of an alternative approach to anaphor resolution – the form-specific multiple-constraints approach which states that a unified ‘monolithic’ notion of salience is insufficient for determining an appropriate reference form, and that individual anaphoric forms carry their own additional constraints on the conditions of their felicitous use.

As we have reviewed earlier, referential forms differ in how specific or informative they are. For example, pronouns and null forms are the shortest anaphors available, they provide the least information, and therefore they tend to be the most restricted in their interpretation, referring to the most accessible entity (Gundel et al., 1993). More fully specified forms such as names, definite and indefinite referring expressions provide more information and can be used to refer to increasingly less accessible entities. Although it is intuitively appealing to define anaphor usage in terms of salience or accessibility, there are several issues with existing proposals. First, it is difficult to define what exactly salience or accessibility is. Second, it is difficult to establish the linking hypothesis between any given salience level and the
reference form used. The notion of accessibility or salience is intrinsically continuous as suggested by Givon (1983) while accounts of salience appeal to discrete cognitive states. And, as noted earlier, salience is often used to explain the preferred use of different referential forms, while the preferred form is used as a diagnostic for salience.

**Defining Salience**

It is notoriously difficult to define accessibility or salience. One approach is to define it in terms of neural ensemble activity, but the relationship with mental discourse representations would be tenuous at best, as the current level of neuroscience does not allow to differentiate between entities of the same category or type, for example, “a beautiful rose” and “a flowering lotus”. Another possibility is to use visual salience of objects and entities as a guide to their accessibility in the discourse model. However, this would significantly limit the domain of applicability of such a definition. Imageability ratings have been suggested as a proxy for conceptual accessibility of entities (Bock & Warren, 1985; Eberhard, 1999), but this is again limited to a restricted domain which ignores any abstract or mental concepts.

For our purposes, the most useful definition of mental salience is in terms of the ease of retrievability of information from memory. This approach also allows us to deal with the problem of salience rankings where there is no clear link between the salience of the referent and the anaphor’s informativity. If an entity is already represented in working memory, its retrieval cost is low, it is readily accessible for
mental processing, and can be referenced with a non-informative anaphor. If, however, the entity has not been mentioned, seen or referenced, its retrieval costs will be higher and it will therefore be harder to access and integrate with the current discourse. The information contained in the anaphor must be greater in this case to assist in locating the intended referent. This approach was first proposed in the Information Load Hypothesis by Almor (1999) – “identifying the antecedent is a matter of reactivating its representation in working memory.”

**Information Load Theory**

The Information Load Theory proposed by Almor (1999) aims to characterize the circumstances in which one form or referring expression is used instead of another in functional terms. It is one of two main theories attempting to provide an integrated theory of NP anaphora processing.

It is generally not clear why the following principle holds – the more salient the referent is, the less information is contained in the anaphoric expression. Definite NPs are considered to have higher processing costs than pronouns. Establishing reference to a highly accessible entity only requires a low-cost referring expression, whereas establishing reference to an entity that is not highly accessible requires an expression with a higher cost. Almor’s theory assumes that referential forms differ from each other in terms of the processing cost. The theory establishes a tradeoff between the processing cost and the functionality of an anaphor, where functionality is defined in terms of whether the use of this particular form is justified – listeners
expect speakers to choose anaphors whose processing cost is balanced by discourse function and are slowed down when this expectation is violated. Assuming that the language processing system strives to operate efficiently, any high cost anaphor use must be functionally justified – either to activate a low salience referent or to introduce extra information into discourse. This cost is defined in terms of conceptual representation – the semantic distance between the preferred referent of the anaphor and the representation of the actual antecedent (Almor, 2000).

Almor argues that the ease of processing NP anaphors can be described by the interaction of three factors: discourse focus, the amount of new information added by the anaphor, and the informational load of the anaphor-antecedent pair. Identifying the antecedent is a matter of reactivating its representation in working memory. The weaker this representation is, the more overlap would be required between the anaphor and the antecedent to reactivate its representation.

**Centering Theory**

Another major theory of reference processing is Centering Theory (Grosz, Joshi, Weinstein, 1995). Centering was developed within computational linguistics and was formulated as a model of the relationship between the attentional state, the form of the referring expression, and the coherence of an utterance within a discourse segment. Its aim was to compute the antecedent of a pronoun based on a simple algorithm, which was later proposed as a hypothesis for how people resolve reference. The basic principle of this approach is that all arguments in an utterance
are ranked according to their grammatical function (Subject>Object>Object2). These ranked elements form the “forward-looking centers” set which serves as potential links to the next utterance by providing possible antecedents for any subsequent anaphors. The highest ranked set member is called the “preferred center” and is considered the most likely antecedent for the next anaphor. Another member of the set is called the “backward-looking center” which provides a link to the previous utterance by referring back to something in the previous discourse (Gordon, Grosz & Gilliom, 1993). The strongest empirical evidence for centering comes from the repeated name penalty – a slowdown in processing when the same name rather than a pronoun is used to refer to an entity that was the focus of the preceding sentence.

This simplified view of reference processing comes from the need to resolve reference through easy computations. However, due to its oversimplification it is prone to errors and inconsistencies. It has been difficult to determine how exactly the arguments should be ranked and Centering theorists acknowledge that other factors, such as the thematic role of an argument, may influence the ranking of forward-looking centers (Grosz et al., 1995; Brennan et al., 1987). Centering makes predictions primarily about pronoun use. However, it is unclear what governs the use of all other anaphoric forms. Also, while the forward-looking centers are ranked in prominence or salience, producing a graded set, predictions are limited to the extremes of this set and the rules governing the rest of the continuum are underspecified.
The restriction of Centering to reference resolution for pronouns within a discourse segment causes several problems for the theory. The theory has difficulty handling topic shifts despite repeated attempts to account for them. Also, recent work has shown that listeners perceive segment boundaries at various levels of granularity (the problem of embedded segments) (Walker, 1998). Some centering theorists have appealed to discourse segments and segment popping – whereby if a topic change is detected, the previous forward-looking set is saved until the discourse comes back to it, and a new set is formed. Although practical from the computational perspective, this account is problematic because it does not explicitly specify the memory mechanisms by which segment popping can be achieved, i.e., how a full representation of the previous segment can be preserved and then retrieved from memory. Another issue is that centers are often continued over discourse segment boundaries with pronominal referring expressions whose form is identical to those that occur within a discourse segment. In addition, even for utterances within a discourse segment, there are strong contrasts between utterances whose adjacent utterance within a segment is hierarchically recent and those whose adjacent utterance within a segment is linearly recent. Also, Centering had been developed primarily to explain pronoun use and distribution and as such does not account well for the patterns of use of other referential forms. The theory suggests that there is only one highest-ranked forward-looking center in each utterance, which can then be referred to with a pronoun. However, it is entirely possible to have situations where
there are two or more plausible candidates for pronominal reference (Garnham, 2008).

Despite attempts to resolve these problems and integrate Centering with a model of global discourse processing (Walker et al., 1998), there is another serious drawback, which is based on the central premise of Centering Theory and therefore cannot be easily addressed. Many theories of reference processing have stipulated strictly linguistic principles on the basis of distributional analyses (Gordon, Grosz, Gilliom, 1993). However, identifying a distributional pattern does not explain why this pattern exists in the first place. As Almor (2000) argues, “an appropriate explanation of processing of anaphoric forms must be based both on the linguistic principles that can be derived from distributional analyses of anaphoric expressions and on psychological principles of conceptual representation.” However, Centering Theory (Grosz, Joshi, Weinstein, 1995) attempts to explain anaphor distribution and processing solely on the basis of distribution-derived generalizations that do not relate directly to any psychological mechanisms.

**Other Approaches**

Despite being both linguistically and psychologically motivated, Almor’s theory also suffers from some drawbacks. It fails to consider the role of possible alternative antecedents for an anaphor by considering only the actual antecedent in calculating the processing cost of the anaphor. However, the question of what other alternatives are considered depends, among other things, on the type of the referring
expression, prior discourse context and individual salience of discourse entities. Garnham and Cowles (2008) also argue that the Information Load Theory does not fully consider the discourse role of the anaphor itself. In his discussion of anaphors that introduce new information Almor seems to assume that this new information is more specific than that provided by the antecedent, for example in the antecedent-anaphor sequence: “… the bird … the robin …” (Almor, 1999), which is debatable. The effects of the relative specificity of antecedent and anaphor have been investigated by Garrod and Sanford (1977) and Garnham (1989), although they produced differing results. Garnham and Cowles (2008) argue that the information content of the anaphor must be considered not only in relation to its antecedent, but also in relation to the discourse context and information about both the referent and all possible antecedents.

In response to the problems with the models discussed, Garnham & Cowles (2008) propose a new JANUS model of noun-phrase (NP) anaphor processing that builds on the existing research and provides an improvement over the Centering Theory (Grosz et al., 1995) and the Information Load Theory (Almor, 1999) by offering a more integrated account of NP anaphor processing.

The theory is essentially a cross between Centering and the Information Load Theory, attempting to address the drawbacks of both theories. As such, it is based on the same general assumption – namely, that the form and content of an anaphor can be explained by the need to establish its referent and the possibility of adding new information in the process. In its current formulation, JANUS adopts a strong
functional perspective, suggesting that the form of a referring expression should reflect its purpose. The model’s basic claim is that to understand how coreferential NP anaphors work, their backward- and forward-looking functions need to be considered. The backward looking functions are primarily concerned with identifying the antecedent (and referent) of the anaphor from among a set of actually considered possibilities. The forward-looking functions are related to the upcoming information about the referent within the anaphor’s clause and in the upcoming discourse. One possible example would be to signal a thematic shift. JANUS, like the Information Load Theory, acknowledges the important role working memory plays in the process of reference resolution and that memory limitations in processing and storage affect the anaphor use and interpretation. The model still needs to be put to an empirical test to identify the validity of its predictions.

There are other plausible approaches that suggest a plausible mechanism for the link between an anaphor and the referential form referring to it. These developments are also consistent with our proposal on the relationship between memory retrieval and salience.

One approach is the Harmonic Alignment formalized by Aissen (1999). Harmonic Alignment assumes that certain types of entities with specific salience characteristics tend to align with specific syntactic structures – in other words, salience, syntactic and thematic hierarchies align with each other. This approach suggests that the primary influence on lexical choices would come not from availability of an entity in memory, but for structural reasons. The intuition shared by
many linguists and adopted by Aissen (1999) is that agents make better subjects than
patients do. Aissen (1999) proposes that semantically ‘active’ (proto-agent)
arguments align with the most prominent syntactic argument positions – they are
attracted to the structural ‘peak’ of the clause, the Subject; and semantically
‘inactive’ (proto-patient) arguments align with the least prominent syntactic positions
– attracted to non-Subject positions. The ultimate reasons for this alignment lie in the
pragmatics of discourse and people’s cognitive biases (Givon, 1983; Ariel 1991;
Gundel et al., 1993; Warren & Gibson, 2002). The modeling of detailed effects of
harmonic alignment on grammars has been expanded to stochastic Optimality
Theory by Bresnan, Dingare and Manning (2001) to capture gradient detail patterns.

Another approach is the Memory Facilitation Hypothesis by Hofmeister
(2007), which proposes that linguistic elements that encode more information
(lexical, semantic, syntactic, etc.) facilitate their subsequent retrieval from memory.
The hypothesis is based on the concept of distinctiveness – as more features are
encoded, the corresponding representation stored in memory stands out more from
any competitors. Hofmeister’s hypothesis shares some common assumptions with
the Information Load Hypothesis (Almor, 1999) – both are tied to the issue of the
retrievability of an entity’s representation from memory.
Overview of Thesis

The goal of this dissertation will be to address both methodological and theoretical questions of real-time discourse processing to understand how listeners coordinate and structure information in online speech. We will contrast the unified salience based accounts of reference resolution (Gundel et al., 1993) with the form-specific multiple-constraint framework (Kaiser, 2003) in which multiple factors, in addition to discourse salience, mediate the mapping between referential forms and referents, with different referential forms being more sensitive to some factors than others. We investigate mechanisms of reference resolution and its relationship to the relative salience of the potential referents by examining participants’ eye movements, choices and acceptability judgments of changes in scenes as they follow instructions to manipulate objects presented in a visual display. We will focus mainly on three specific referential forms (“that” N, “the” N and “a” N) in order to (a) evaluate the proposal that the choice of referential forms is based on the relative salience of the referent and (b) develop a methodology that tests the salience of a discourse referent.

As we have discussed above, many theories of how different anaphoric forms are generated and interpreted appeal to salience as a central concept. With this experimental approach we aim to investigate the mechanism of discourse reference resolution at various levels of referent salience. Different referential forms are assumed to be most effective at activating referential antecedents under different discourse conditions. We manipulate the cognitive status of discourse entities by
varying the type of discourse material that intervenes between the antecedent and the subsequent referring expression and assess the accessibility of the referent by examining how listeners identify the referent of different anaphoric forms. We also examine which alternative referents are considered during the reference resolution process.

We focus on three different anaphors assumed to refer to entities of progressively lower salience: the demonstrative (that N), definite reference (the N), and indefinite reference (a N) as exemplified by the discourse: “Click on the heart above the lemon. Now click on that heart/the heart/a heart”. These three forms are categorized into different levels of salience or accessibility on the Givenness Hierarchy (Gundel et al., 1993). We will examine the behavior of these different anaphors under varying discourse conditions to see whether the same referential form would behave differently under different circumstances. We also investigate whether these three forms carry different licensing constraints on the conditions of their felicitous use and what these form-specific constraints may be.

In Experiments 1, 2 and 3 we present listeners with a visual display containing multiple objects, enter two of them into a discourse model in an initial “setup” sentence (e.g., “Click on the heart above the lemon”) and introduce various amount of intervening discourse material between the setup sentence and a sentence containing the anaphor. In these studies the anaphors are consistent with either the object referred to by the antecedent or with another identical object. Experiments 1 and 2 investigate the impact that initial salience of a referent has on the preferences
of the referring expression this referent is subsequently referred to with and whether changing discourse affects these preferences. Experiment 3 examines the conditions of felicitous use for the indefinite referential form. Finally, in Experiment 4 we investigate the same questions as in Experiments 1-3 using an acceptability rating instead of having the participants actively choose the referent. Additionally, we attempt to measure the salience of a potential referent directly by measuring the listeners’ sensitivity to visual changes in the display, adopting the influential change blindness paradigm (Rensink et al., 1997). This approach aims to resolve the circularity of salience definition that we have discussed earlier.
Chapter 2 – Experiments on Dynamic Changes in Discourse Salience

As we outlined in the previous chapter, the most comprehensive salience-based approach to anaphoric forms is the Givenness Hierarchy (GH) proposed in Gundel et al. (1993, see also Givón 1983, Ariel 1990). The Givenness Hierarchy proposes that different referential forms refer to entities with different levels of “salience”, “accessibility” or “activation”. The assumption behind the gradation of salience levels is that each referential form is preferentially used for discourse referents with a specific degree of salience. The most reduced referring expressions, such as null anaphors in so-called “pro drop” languages where subjects can be omitted, and pronouns, can only be used to refer to the most salient referents/discourse entities. Fuller linguistic forms are used for less salient referents.

The Givenness Hierarchy is reproduced in Table 2.1. The headings represent the cognitive states proposed by Gundel et al. (1993) for English, with the states decreasing in salience from left to right. The linguistic forms are listed below the heading. When there is more than one form listed under a heading, the forms are ordered from most accessible to least accessible. Each referential form is associated with a minimal level of salience required for use of that form. A form lower on the hierarchy can be used to refer to a more accessible form, but with some reduction in felicity.
Table 2.1: Givenness Hierarchy and Associated Anaphoric Forms in English

<table>
<thead>
<tr>
<th>in focus</th>
<th>activated</th>
<th>familiar</th>
<th>uniquely identifiable</th>
<th>referential</th>
<th>type identifiable</th>
</tr>
</thead>
<tbody>
<tr>
<td>it</td>
<td>that, this</td>
<td>that N</td>
<td>the N</td>
<td>indefinite</td>
<td>a N</td>
</tr>
</tbody>
</table>

This N

The evidence in support of the Givenness Hierarchy comes primarily from linguistic intuitions and corpus analyses. Relatively little experimental work has been specifically aimed at investigating the salience hypothesis. The goal of our experiments is to develop an experimental approach in which we can rigorously evaluate the salience hypothesis. The experiments presented in this chapter make use of a simple paradigm in which we manipulate the form of the referring expression and the prima facie salience of the referent of the referring expression through simple linguistic and discourse manipulations.

We selected forms from three different salience levels in the Givenness Hierarchy that can be used to refer to the same type of referent. In order of diminishing salience, these are that N (familiar), the N (uniquely identifiable), and a N (type identifiable). The definitions of these states proposed by Gundel et al. (1993) are as follows:

**Familiar:** The addressee is able to uniquely identify the intended referent because there is already a representation of it in the addressee’s memory. The referent has either been mentioned before or the addressee is aware of its relevance to discourse.
**Uniquely identifiable:** The addressee can identify the intended referent on the basis of the nominal alone. The referent typically either has been mentioned at some point or is otherwise unique.

**Type identifiable:** The addressee is able to access a representation of the type of object described by the expression – the referent typically has not been mentioned in the current discourse or is non-unique. This form would normally be used to introduce new referents (Warren & Gibson, 2002).

In Experiments 1 and 2 participants viewed a display, which consisted of a 4 by 4 grid, with 16 black and white pictures of objects (4 x 4). Some objects were doubled (two tokens of the same picture), and some pictures had names that shared the same initial sounds, e.g., “heart” and “harp”. We will refer to these as “cohort competitors”. This terminology is adopted from the classic Cohort Model (Marslen-Wilson, 1987), which proposes that during spoken word recognition, words sharing the same initial sounds become activated together as thus form a cohort. We included cohorts because they provide a subtle test of which potential referents are considered as a referring expression unfolds. In particular, evidence that a referent is being considered comes from an increase in looks to that referent when its name is temporarily compatible with the unfolding referring expression. Participants then heard a sequence of two or three instructions. The first instruction was of the form, “Click on the X above the Y” and will be referred to as the discourse setup sentence. The test sentence always came last and referred to one of the entities mentioned in the first instruction, using one of the three forms: *that N*, *then N* or *an N*. 
Discourse status was manipulated in two ways. First, we varied whether the referent of the test sentence was initially introduced as the Theme (X) or the Location (Y). In an utterance of the form “Click on the X above the Y”, the Theme is hypothesized to be a more salient entity than the Location. The Theme and Location objects were always doubled in the display. Second we manipulated whether or not another instruction intervened between the set-up sentence and the test sentence.

There were three possibilities. In the **blank** condition, the test sentence immediately followed the setup sentence. In the **look at** condition, participants were instructed to look at a cross. Unlike the other objects in the display, the cross could not be clicked on, and was always used in “look at” instructions, which were also used in eye-tracker calibration checks. This allowed us to have an intervening instruction expression that did not introduce a new discourse referent. In the **click on** condition, participants were instructed to click on another object. Thus there was an intervening sentence that introduced a new discourse entity.

The experimental setup is summarized below:

– The **discourse setup** instruction introduced the Theme and the Location – varying the initial salience of the referent.

– The **intervening** instruction – manipulating the change of salience of the discourse entities established by the setup sentence {**blank**, **look**, **click** conditions}

– The **test** instruction manipulated the referential form used to refer to one of the entities introduced as the Theme or the Location.
Figure 2.1: Sample Display, Experiments 1 & 2.
Examples of the different conditions are presented in Table 2.2 for the sample display in Figure 2.1. In sentence (1), the heart above the lemon in the display is the referent introduced by the Theme and the lemon below the heart is the referent introduced by the Location modifier.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>Sample Discourse, Experiment 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>Click on the heart above the lemon.</td>
</tr>
<tr>
<td>2</td>
<td>a</td>
<td>&lt; blank &gt;</td>
</tr>
<tr>
<td>2</td>
<td>b</td>
<td>Now look at the cross.</td>
</tr>
<tr>
<td>2</td>
<td>c</td>
<td>Now click on the broom.</td>
</tr>
<tr>
<td>3</td>
<td>a</td>
<td>Now click on that heart.</td>
</tr>
<tr>
<td>3</td>
<td>b</td>
<td>Now click on the heart.</td>
</tr>
<tr>
<td>3</td>
<td>c</td>
<td>Now click on a heart.</td>
</tr>
</tbody>
</table>

**Experiment 1** – Theme (high salience)

Experiment 1 seeks to determine whether there is variation in the goodness of match of a high-salience discourse entity for different referential forms in varying discourse conditions. By manipulating the discourse change through introduction of various degrees of intervening material between the initial discourse setup and the test reference, we aim to examine how the process of reference resolution changes for the three referential forms we are interested in. In this experiment the noun in the referential sentence (*that N, the N or a(n) N*) was always the same noun as the Theme in the setup instruction. Our working assumption is that as the discourse representations change, so the preferred way of referring to the Theme should change.
as well. We also assumed that with more intervening material (either in terms of discourse distance from the initial mention, or in terms of new discourse entities introduced into the model), the initial mental representation of the Theme will become less salient, and consequently it should be less accessible as the referent of a referring expression.

We monitored eye movements using the visual world paradigm (Cooper, 1974; Tanenhaus, Spivey-Knowlton, Eberhard & Sedivy, 1995) as participants followed pre-recorded spoken instructions to look at or click on the pictures in the display. Each trial consisted of two or three sentences: (a) a discourse setup sentence which introduced the Theme and the Location; (b) an intervening sentence, which manipulated whether linguistic material or a discourse entity intervened between the antecedent and the anaphor; and (c) a test sentence, which manipulated the referential form used to refer to the Theme or the Location.

Examples of these different conditions are presented in Table 2.2 (1)–(3) for the sample display in Figure 2.1. In sentence (1), the heart above the lemon in the display is the Theme and the lemon below the heart is the Location. The discourse referents are introduced by the first instruction (1). It establishes two entities in the discourse model in a relationship to each other – the Target (Heart) and the Location (Lemon). To manipulate the salience of entities in this initial discourse representation, we used one of three types of intervening material (2)(a-c):
(2a) ‘Blank’ condition – no intervening sentence, the final test sentence follows immediately after the first one. This does not change the initial salience distribution in any way and serves as a baseline.

(2b) ‘Look’ condition – an intervening sentence that presumably does not introduce a new discourse entity. The instruction asks the participant to perform an action unrelated to the prior discourse setup. At the start of the experiment, participants are instructed that sometimes they will be asked to “look at the cross” for eye-tracker calibration purposes. When the eye gaze lands on the picture of a cross that is embedded within each display (Figure 2.1), the program automatically advances the trial to the final sentence. This makes this condition functionally different from the task of clicking on objects, and the cross is also not a part of the possible object set – it is never clicked on. This, presumably, creates a new discourse segment, thereby allowing the participant to isolate and preserve the initial salience setup (Walker, Joshi, & Prince, 1998).

(2c) ‘Click’ condition – an intervening sentence that introduces a new discourse entity while maintaining coherence of the discourse segment and the task, thereby changing the mental status and salience of the representation of the entities within the discourse model.

It is important to note that ‘look’ and ‘click’ conditions are equated for the linguistic distance and time between the initial mention and the subsequent reference, with only the ‘click’ condition introducing a new discourse entity.
The final test sentence refers to the Theme entity (*Heart*) of the initial sentence using three referential forms: (3a) – *that* N, (3b) – *the* N, (3c) – *a* N. We use the timing and distribution of looks and choices to the Theme (the heart above the lemon) and to the other heart to infer how the discourse status of the theme and the form of the referring expression affect reference resolution.

**Predictions**

The implicational nature of the Givenness Hierarchy (Gundel et al., 1993) allows us to make specific predictions of the behavior of the three selected referential forms. The model suggests that each discourse status entails all lower statuses, but not vice versa. For example, an entity that is in focus, is necessarily also activated, familiar, uniquely identifiable, referential and type identifiable. However, not all uniquely identifiable entities are familiar or in focus. Therefore, referents that would be a good match for ‘*that* N’ expression, should also match well the ‘*a* N’ form, but not vice versa.

Potential form-specific licensing conditions on the different anaphors due to different definiteness definitions also enter into our predictions. While definites and demonstratives usually refer to a real, specific thing or concept, indefinites may not refer to a specific thing, but rather introduce a general description. Also, indefinites are usually considered to introduce new information into the discourse.

We also predict an increased consideration of potential competitor items when there is more intervening material (2) between the setup sentence (1) and the
test sentence (3). More intervening material should result in an increased proportion of looks to the cohort items (e.g., the harp) and the alternative referent (the other heart).

The predictions for the different kinds of referring expressions are outlined below:

**That heart** – In all three conditions (2a, 2b, 2c), we expect participants’ final referent choices to show a strong preference for the already-mentioned heart, since ‘that N’ is hypothesized to require a previously mentioned, highly salient referent. For the different intervening material conditions, we predict that there should be relatively few looks to either the alternative referent or the cohorts in the ‘blank’ and ‘look’ conditions (2a, 2b), since ‘that N’ can only be used felicitously to refer to mentioned items. In the ‘click’ condition (2c), we expect to see slightly delayed looks to the referent, with perhaps some looks to the alternative referent. We do not expect to see looks to cohorts, due to ‘that N’ being restricted to previously mentioned referents.

**The heart** – In all three intervening conditions, we predict that participants’ final choices will indicate a preference for the Theme, the mentioned heart, over the non-mentioned heart. This prediction is based in part on an assumption that “the” is preferentially taken to refer to a previously established discourse referent rather than introduce a new discourse referent. However, in all three intervening material conditions, we expect to see competition from the unique cohort (the harp), since ‘the N’ can also be used to introduce a new referent, that is to refer to a linguistically
unmentioned but visually unique entity in the display. We further predict that the proportion of fixations to alternative referents and to cohort items will increase as the amount of intervening material increases (2a<2b<2c).

A heart – Here we expect to see a preference for, and more looks to, the non-mentioned item (another heart), with some consideration of the mentioned referent. However, we predict that this preference will diminish with the increase in the amount of intervening material (2a>2c), since more intervening material should make it easier to interpret ‘a heart’ as referring to the already-mentioned heart since it should become less salient. This should lead to a higher proportion of fixations to the mentioned entities.

**Methods**

**Participants**

Eighteen male and female University of Rochester undergraduate students who volunteered to participate in the experiment were paid $7.50 for their time. All participants were native speakers of English, had normal or corrected to normal vision and no reported hearing difficulties.

**Materials**

All auditory stimuli were pre-recorded in a noise-proof laboratory on a DAT tape and then digitized at 44kHz. All trials were recorded individually with all three
sentences following after each other to avoid any discontinuity of pitch, voice gaps or irregularities. The speaker was instructed to use normal intonation and avoid placing emphasis on any part of any sentence to avoid biased interpretations.

We presented potential referents in a visual display – a grid of 16 objects (4 x 4) presented on the screen (see Figure 2.1). The grid was centered on the computer screen, with at least 50 pixels of white space around the edges of the screen. The display resolution used was 1280 x 1024 pixels. Each grid-cell was 200 x 200 pixels in size and subtended approximately 5 degrees of visual angle horizontally and 4.8 degrees vertically. There was a 40 pixel distance between the cells horizontally and 30 pixels vertically. All items were black and white line drawings of familiar objects. All objects were 150 pixels on the longest dimension and had at least 80 pixels of white space between them. Items were approximately 3.7 degrees of visual angle in diameter at a 70 cm viewing distance.

As mentioned earlier, we will refer to the two items mentioned in the initial sentence (Heart and Lemon) as the Theme and Location respectively. The Theme is the item the participant is instructed to click on in the initial sentence, and the Location is the item that identifies the intended Theme. Apart from these two items, the display contains possible phonological competitors for both the theme object and the location object.

– for Heart (the ‘Theme item’):
(i) another identical Heart above another item (potential second target location)
(ii) a unique cohort – Harp (which can be felicitously referred to with the definite article, e.g., ‘the harp’)

– for Lemon (the ‘Location item’):

(i) another identical Lemon

(ii) a unique cohort – Leopard (which can be felicitously referred to with the definite article, i.e. ‘the leopard’)

(iii) cohort pair – two identical Legs (each of which can be felicitously referred to with an indefinite article, i.e. ‘a leg’)

These items were then intermixed with a set of distracters, which included: the Location of the second Heart (the item that the second Heart is above) and its identical pair, a cohort pair for this location item, 3 unrelated objects, a cross (“for calibration purposes”). Overall, there were 16 pictures in each display.

All pictures were selected so that their names were not phonologically similar (except within cohort pairs or triplets), and they were not visually similar, semantically related or associated with each other. For this purpose, we created a database of objects containing over 600 pictures.

All items are reused three times during the course of the experiment. In particular, all cohort items are used once in each of these roles: Target of an instruction, Location item in the instruction and an unrelated Distracter. Thus, each
mentioned item is mentioned twice – once as the Theme to be clicked on, and once as a Location.

Previous pilot experiments revealed a central fixation bias for the displays. Each trial begins with the participant looking at a centrally positioned cross for drift correction. Thereby the central area around the cross (2 x 2 cells) attracts most of the initial attention, and the objects positioned in these grid slots are more visually salient at the onset of the trial. Thus, the first saccade launched after the disappearance of the cross has the highest probability of landing on one of these objects.

We took advantage of this bias by positioning most relevant competitors in this central area, e.g., one of the lemons, one of the legs, the unique leopard and the unique harp. The Theme is never located in this central area, thus ensuring that the participants have to explore the display to find it. The last remaining leg is positioned immediately next to the Theme of the initial instruction (heart) to make sure participants notice that there are two legs in the display – which is important for the indefinite conditions. This maximizes the participant’s awareness of the relevant items and their locations. The rest of the items were randomly positioned in the display so that no two tokens of the same item were located immediately next to each other.

The items appearing in the intervening material sentences – namely the cross and one of the distracters – were positioned at a roughly equidistant location from the target of the discourse setup sentence to equate the visual distance the eye travels
when the participant looks away from the discourse theme. This distance was equated to the extent that the grid setup of the display allows – to within a one-cell difference in position.

The experimental trials described above always contain the target of the first sentence, either in the top row of the 4x4 grid or in the left/right columns, but never in the central area. The potential competitors were distributed between the central area and the periphery. The final test instruction always referred to the Theme of the initial sentence using varying referential forms.

The filler trials contained the following conditions to disguise the experimental manipulations:

– The competitors were located in the periphery and the objects mentioned in the discourse setup sentence located in the central area.
– The last instruction referred to an unrelated item in the periphery while the competitors were located in the central area.
– The competitors were located peripherally and the unrelated items were located centrally with the last instruction referring to either the competitors (peripheral) or the unrelated distracters (central).

All trials were randomized so that no more than two consecutive trials had either the same intervening sentence type, the same referential form, the same location of the Theme, the same item used as the Theme or the Location, or the same type of filler condition. There were a total of 126 trials with 45 experimental trials.
which resulted in five data points for each of the nine conditions: three intervening material conditions x three referential forms.

**Procedure**

An SRI EyeLink II head-band mounted eye-tracker was used to monitor participants’ eye movements, sampling at a rate of 250Hz. Whenever possible, attempts were made to track both the pupil and the corneal reflection of both eyes to improve accuracy. The calibration procedure used a 9-point calibration grid, and a drift correction was performed before every trial using the same picture of a cross as the one embedded in the displays and referred to in the ‘look at the cross’ instruction. Experimental stimuli were presented using custom software written in Visual C++. Pre-recorded auditory stimuli were delivered through the Sennheiser HD 570s headphones.

The participants’ task was to follow the spoken instructions to click on the items in the display. At the beginning of the experiment the participants were informed that if they took too long to respond, the trial would time out. However, they were not asked to respond as quickly as possible, nor were they aware of the time-out delay (3 sec). All participants were seated comfortably approximately 70cm from the computer screen with the monitor at eye level. Thus each object subtended approximately 3.7 degrees of visual angle.

Each trial started with the centrally positioned cross on the screen which the participants were required to fixate and then click the computer spacebar to perform
the drift correction. 100 ms later the display with the pictures appeared followed by
the onset of an auditory instruction (the first/discourse setup sentence). The onset of
each instruction was 100 ms after the onset of the display. After each instruction, the
program paused until the participant made a response with the computer mouse, with
a timeout after 3 seconds. At the end of each trial, the program paused until a
successful drift correction was performed, allowing participants an opportunity to
rest if necessary. The experiment lasted approximately 50 minutes excluding the
initial explanations, setup and calibration of the eye-tracker.

**Data Collection and Analysis**

The data were collected within the eye-tracker-generated EDF data files,
error corrected and then processed using custom made data-analysis software written
in Microsoft Access. Further analysis involved determining the proportion of
fixations to each object in 4ms time slices. An eye movement was considered to have
landed on an object if the participant fixated anywhere within the item’s grid cell.

The eye-movement data provide incremental on-line information about the
reference resolution process, while the participant’s chosen referent (what they chose
to click on) indicates the final choice of the referent. Only trials that did not timeout,
did not have any errors, and in which the participants clicked on a plausible object
were included in the analyses (less than 1% of data was excluded).
Post-experimental debriefing of participants established that none became aware of the experimental manipulations, and thus all participants, who have successfully completed the experiment, were included in the final analysis.
Figure 2.2: Referent Choices, Experiment 1. (a) Top panel: Percent of mentioned referent choices across different intervening material conditions for all three referential forms (that N, the N, a(n) N). (b) Bottom Panel: Percent of choices for mentioned and non-mentioned referents in the indefinite reference form condition. Error bars represent SEM.
Results

Choice of referent

Figure 2.2a presents the proportion of the mentioned item choices when the theme was selected as the target of the test sentence and clicked on. It displays proportions for the different referential conditions (that N, the N and a N) across the three intervening material conditions (‘blank’, ‘look’ and ‘click’). A repeated measures ANOVA was performed on the quasi-empirical logit-transformed data to eliminate typical errors associated with ANOVA analysis of categorical data (Agresti, 2002; Jaeger, in press). The data transformation was defined as $X' = \ln((((X-0.5)*0.99)+0.5)/(1-((X-0.5)*0.99)+0.5)))$.

The analysis revealed significant main effects of referential form, $F(1,17)=38.869$, $p < 0.001$, and intervening material, $F(1,17)=47.010$, $p < 0.001$, as well as a significant interaction, $F(1,17)=6.382$, $p = 0.0217$. Specifically, the ‘blank’ and ‘look’ conditions are significantly different from the ‘click’ condition, but not from each other. The indefinite referential form is significantly different from both the definites and demonstratives, but they are not different from each other. Note that in the ‘blank’ and ‘look’ conditions there is no effect of the referential form used – the proportions of mentioned referent choices are not significantly different from each other. However, in the ‘click’ condition, the indefinite form elicits a significantly smaller proportion of choices of the mentioned referent than the other two forms.
Overall, there is a preference for the mentioned referent even in the indefinite form condition.

A more detailed examination of the proportion of choices for the indefinite referential form condition ‘a N’ (Figure 2.2b), comparing mentioned and non-mentioned referents, confirms a preference for the mentioned discourse active entity across all intervening material conditions. A repeated measures ANOVA on the transformed data shows a main effect of mention, $F(1,17)= 60.065, p < 0.001$, but no effect of intervening material, $F(1,17)= 1.827, p = 0.194$. However, this preference is significantly weakened in the ‘click’ condition, giving us a strong interaction, $F(1,17)=18.060, p < 0.001$. The mentioned referent choice proportions are significantly greater for ‘blank’ and ‘look’ conditions than for ‘click’ condition, and the non-mentioned referent choice proportions are significantly smaller for the same comparison.

Eye-movements

A further analysis of the participants’ eye movements allows for an in-depth look at the real-time reference resolution process and allows us to examine the timeline of when this preference for the mentioned entity arises.

Eye movements during the initial discourse setup sentence (Figure 2.3) reveal both the central fixation preference bias described before and the parallel activation of the two possible targets (Hearts) with the correct one (above a Lemon) winning over time.
Figure 2.3: Fixation Proportions during Discourse Setup, Experiments 1 & 2. Proportions of fixations over time to the objects on display during the discourse setup sentence. The zero point is the moment of display onset. The auditory stimuli onset is around 410 ms later. The vertical bars indicate moments of sentence onset, the Theme noun onset, and the Goal noun onset.
Figure 2.4: Fixation Proportions on the Mentioned Referent. Proportions of fixations over time to the mentioned referent across different intervening material conditions (blank, look, click), grouped by the referential form used in the test instruction (that N, the N, a(n) N). Only trials in which participants were not looking at the mentioned referent at the onset of the determiner are included.
Figure 2.5: Fixation Proportions on the Non-Mentioned Referent. Proportions of fixations over time to the non-mentioned referent across different intervening material conditions (blank, look, click) grouped by the referential form used in the test instruction (that N, the N, a(n) N). Only trials in which participants were not looking at the non-mentioned referent at the onset of the determiner are included.
The four centrally positioned competitor objects capture eye fixations at the onset of the display with the proportion of fixations to these items starting to rise about 150-200 ms from display onset. Given prior studies of saccadic planning, signal driven fixations are expected to begin approximately 200ms after the onset of the stimulus (Matin, Shao, & Boff, 1993)). These central items are the second lemon (goal double), one of the legs (goal non-unique cohort), the leopard (goal unique cohort), and the harp (target unique cohort). The proportion of fixations to the two potential targets (hearts) start rising simultaneously around the time of onset of the target word (980 ms), followed closely with a small delay by the fixations on the two Locations defining the Target’ position. The point of separation of the Target from its pair occurs about 100 ms after the onset of the location word (1760 ms), which serves in this case as a uniqueness point. Overall, this demonstrates that the participants examine the display in line with our expectations, identifying and locating all the relevant items. Thus, the discourse setup sentence achieves its purpose, and the discourse model is set up in line with our assumptions.

Figures 2.4 and 2.5 plot the probability of eye fixations on the mentioned referent (Heart that was the target of the first instruction) and the non-mentioned competitor (the other Heart in the display) across the three intervening material conditions. The total probabilities do not sum to 1.0 due to fixations to other items on the screen not being plotted. The zero point represents the onset of the determiner, with each trial aligned separately based on individual onset measurements. In the ‘blank’ condition, participants often do not move their eye gaze off the initial
sentence’s target when nothing intervenes between the initial and the test sentences, which results in a high proportion of fixations to the mentioned target. Since this can potentially introduce errors into the analyses, we have plotted here and will be analyzing only trials when the eye gaze was off the initial sentence’s target at the moment of the determiner onset. Which is why all proportions start from 0 at time 0.

In all three referential forms conditions, we see a rise in proportions of fixations to the target within 200 ms, which is consistent with when we first expect to see signal-driven fixations. We have isolated trials for the different form conditions (‘that N’, ‘the N’, ‘a N’) into separate graphs. Figure 2.4 plots proportion of fixations to the mentioned target of the initial sentence, and Figure 2.5 – to the non-mentioned alternative target. A repeated measures ANOVA was performed on the quasi-empirical logit transformed fixation proportions data over two time windows of 200-600 ms and 600-1000 ms. The window length of 400 ms was chosen to match the average noun length in the test sentence (408 ms). The first window captures fixations from the onset of signal driven eye movements until the participants have heard all or most of the noun – we expect to see differences reflected in the increase of fixations to the target between different conditions in this window. The second window captures the eye movements launched when participants are in the process of picking out a referent – we expect to see differences in terms of the chosen referent here.

For the fixations to the mentioned target (Figure 2.4), the first time window (200-600ms) reveals a significant main effect of referential form, F(1,17)= 16.481, p
< 0.001, and of intervening discourse type, F(1,17)= 35.540, p < 0.001, but no interaction.

Considering individual referential form conditions, in ‘that N’ condition, we find evidence that in the ‘blank’ condition fixation proportions rise faster and higher than in the ‘look’ and ‘click’ conditions (F(1,17)= 7.362, p = 0.0148; F(1,17)= 21.691, p < 0.001), but ‘look’ is not significantly different from ‘click’. In ‘the N’ condition, all discourse type conditions are different from each other – ‘blank’ vs ‘look’ F(1,17)= 8.952, p = 0.008, ‘blank’ vs ‘click’ F(1,17)= 22.100, p < 0.001, ‘look’ vs ‘click’ F(1,17)= 9.992, p = 0.006. Finally, the ‘a N’ condition, is similar to ‘that N’ – the ‘blank’ condition has significantly greater proportion of fixations than the ‘look’ and ‘click’ conditions (F(1,17)= 6.783, p = 0.0185; F(1,17)= 9.216, p = 0.008), but ‘look’ is not significantly different from ‘click’.

The first window (200-600ms) for the non-mentioned target fixations (Figure 2.5) reveals no significant main effects or interactions.

Analysis of the second time window (600-1000ms) for the fixations to the mentioned target (Figure 2.4), reveals, same as for the first window, a significant main effect of referential form, F(1,17)= 29.615, p < 0.001, and of intervening discourse type, F(1,17)= 29.239, p < 0.001, but no interaction.

Turning now to the data for each referential form, the results are similar to those from window one. In ‘that N’ condition, we find evidence that the ‘blank’ condition is different from the ‘look’ and ‘click’ conditions (F(1,17)= 7.109, p = 0.0163; F(1,17)= 9.165, p = 0.008), while ‘look’ is not significantly different from
‘click’. In ‘the N’ condition, the ‘blank’ and ‘look’ conditions are different from the ‘click’ condition (F(1,17)= 11.302, p = 0.004; F(1,17)= 10.517, p = 0.005), while the ‘blank’–‘look’ difference is no longer significant. Finally, the ‘a N’ condition, is now similar to ‘the N’ – the ‘blank’ and ‘look’ conditions are significantly different from the ‘click’ condition (F(1,17)= 11.752, p = 0.003; F(1,17)= 6.423, p = 0.0214), while the ‘blank’–‘look’ difference is no longer significant.

Analysis of the second time window (600-1000ms) for the non-mentioned target fixations (Figure 2.5), reveals significant main effects of referential form, F(1,17)= 39.216, p < 0.001, and of intervening discourse type, F(1,17)= 14.784, p = 0.001, as well as a significant interaction, F(1,17)= 8.949, p = 0.008.

Examining the individual referential form conditions, for ‘that N’ condition, we find no significant difference between intervening discourse type conditions. In ‘the N’ condition, the ‘look’ condition is significantly different from the ‘click’ condition, F(1,17)= 11.072, p = 0.004, while ‘blank’ is only marginally different from ‘click’, F(1,17)= 3.257, p = 0.0889, and the ‘blank’–‘look’ difference is not significant. Finally, for the ‘a N’ condition, we find the same result as for the first window. The ‘blank’ and ‘look’ conditions are significantly different from the ‘click’ condition (F(1,17)= 16.942, p < 0.001; F(1,17)= 11.196, p = 0.004), while the ‘blank’–‘look’ difference is not significant.
**Discussion**

We do not discuss in detail the results of cohort competitors analyses as there were no significant trends. The cohort competitors demonstrated an expected effect of phonological overlap competition (Allopenna, Magnuson & Tanenhaus 1998). The proportion of eye fixations on the cohort items briefly rose at the onset of the target word and then returned to baseline. This effect was not modulated by either the referential form or the type of intervening material.

Analyzing the fixation proportions data, we do not find anything unexpected in the first time window (200-600 ms). The ‘blank’ condition rises to the mentioned target faster (Figure 2.4). In the case of the non-mentioned target (Figure 2.5), the fixation proportions are very low, due to low levels of activation for non-mentioned items, which leads to no significant differences between intervening discourse type conditions.

In the second window (600-1000 ms), for mentioned target proportions (Figure 2.4), the ‘that N’ condition maintains the ‘blank’ advantage over ‘look’ and ‘click’ conditions in terms of rise to target, but in ‘the N’ and ‘a N’ conditions this advantage is reduced and there is no difference between ‘blank’ and ‘look’ conditions. The ‘a N’ condition, however, is changing in this window – it takes longer to reach the target and the level of fixations is lower than that in the other two conditions.
In case of the non-mentioned target (Figure 2.5), we find no differences between conditions for ‘that N’ case – fixation proportions are very low and uniform due to this form’s strong preference for a mentioned antecedent. In ‘the N’ condition we find a trend toward greater differentiation between intervening discourse type conditions, but only ‘look’ vs ‘click’ difference reaches significance. In the indefinite case, however, we find a much greater level of fixation proportions to the non-mentioned referent, in line with the expectation of an indefinite preference for the non-mentioned or non-unique entities.

Overall, the results for the that N and the N conditions conformed to our predictions. For ‘that N’ condition, we find that for all types of intervening material, participants had a strong preference in favor of the mentioned target item with very little competition from the alternative non-mentioned target item or the non-mentioned cohort competitor. For ‘the N’ form, we also find a strong preference for the mentioned target item with a small proportion of fixations to the non-mentioned competitors across all intervening material conditions. The time-window analysis demonstrated a trend for the increased intervening discourse to increase competition from non-mentioned alternative referents.

The results for the indefinite condition ‘a N’, however, did not fully conform to our predictions. We find that the fixation proportions to the non-mentioned referent are higher in the indefinite reference condition compared with ‘that N’ or ‘the N’ cases, in line with the expectation of a preference for the non-mentioned entities. However, according to the predictions, an indefinite referential form should
prefer a non-mentioned entity as a referent and it should do so regardless of the discourse conditions. In contrast we find that the indefinite form strongly prefers a mentioned referent across all intervening material conditions.

A detailed analysis of the fixation proportions in the ‘a N’ condition further confirms the above findings. There is a strong preference for the mentioned referent in the ‘blank’ and ‘look’ conditions with proportions of fixations to the mentioned target reaching 60-70% and 10-15% to the non-mentioned one. However, in the ‘click’ condition, the proportions of fixations are split evenly between mentioned and non-mentioned referents with about 40% each. Furthermore, the second statistical window for the non-mentioned referent was the only one to exhibit a significant interaction between the referential form used and the amount of intervening material. Which suggests that the increase in the intervening discourse is related to the increased competition from the non-mentioned referent. Remember that ‘look’ and ‘click’ conditions are equated for the linguistic distance and time between the initial mention and the subsequent reference, with only the ‘click’ condition introducing a new discourse entity. This demonstrates that introducing discourse relevant intervening material leads to a change of the discourse salience of established entities above and beyond changes due to time and/or presence of linguistic material.

Overall, the introduction of a new discourse entity in the ‘click’ condition decreased the accessibility of the mentioned Theme referent as manifested in the delayed fixations to the referent in the definite and indefinite form conditions, and the increased proportion of fixations to the non-mentioned competitors (i.e. the other
heart) in the indefinite condition. These results are in line with the predictions we generated from the Givenness Hierarchy. However, one surprising aspect of the data was the high proportion of choices of the Theme for the indefinite condition. In addition, because participants were often looking at the Theme in the ‘blank’ condition, there was a baseline difference in looks to the target between the ‘blank’ condition and the ‘look’ and ‘click’ conditions. Therefore, we decided to repeat the same conditions as were used in Experiment 1, but this time the referring expression referred to the Location rather than the Theme. The Location is never clicked on in the setup instruction, therefore participants will be unlikely to be looking at the Location entity during the test sentence in the ‘blank’ condition, thus eliminating the baseline difference. We will postpone discussion of possible explanations for, and implications of, the mentioned bias for the indefinite condition until after we present the results of Experiment 2.

**Experiment 2** – Location (lower salience)

In this experiment we address the concerns from Experiment 1 by exploring the behavior of the same three referential forms when applied to a less salient discourse entity. We do this by focusing on the Location of the discourse setup sentence (1) – the lemon. Because this entity is a Location that defines the position of the Theme (heart), it should be less prominent in the discourse. It is also not a target for fixations following the end of the first instruction. So the baseline in the
‘blank’ condition would not be affected and we should be able to compare the speed and slope of rise to target in all three intervening material conditions more accurately.

**Experimental Setup**

The experimental setup, the display and the procedure remained exactly the same as in Experiment 1 with the exception of auditory instructions (Table 2.3), which now refer to the Location (lemon), not the Theme (heart).

**Table 2.3: Sample Discourse, Experiment 2**

| (1’ | Click on the heart above the lemon. |
| (2’ a) | < blank > |
| (2’ b) | Now look at the cross. |
| (2’ c) | Now click on the broom. |
| (3’ a) | Now click on that lemon. |
| (3’ b) | Now click on the lemon. |
| (3’ c) | Now click on a lemon. |

**Participants**

18 male and female University of Rochester undergraduate students who volunteered to participate in the experiment and were paid $7.50. All participants were native speakers of English, had normal or corrected to normal vision and no reported hearing difficulties.
**Predictions**

The predictions remain the same as in Experiment 1. We expect to replicate our results for the demonstrative and the definite referential forms, and to explore whether the indefinite form will behave as predicted with a less salient non-focused entity (the Location). That is, it should prefer the non-mentioned entity in accordance with the predictions of the Givenness Hierarchy (Gundel et al., 1993).

**Results**

*Choice of referent*

Figure 2.6a presents the proportion of the mentioned item choices when the mentioned location was selected as the referent of the test sentence and clicked on. It displays proportions for the different referential conditions (that N, the N and a N) across the three intervening material conditions (‘blank’, ‘look’ and ‘click’). A repeated measures ANOVA was performed on the data following the same quasi-empirical logit transformation, as in Experiment 1.

Overall the results are similar to those for Experiment 1. The analysis revealed significant main effects of referential form, $F(1,17)=11.440$, $p = 0.004$, and intervening material, $F(1,17)=55.723$, $p < 0.001$, as well as a significant interaction, $F(1,17)=5.040$, $p = 0.0384$. Specifically, the ‘blank’ and ‘look’ conditions have a significantly higher proportion of mentioned target chosen than the ‘click’ condition, but not different from each other. The indefinite referential form demonstrates a
significantly smaller proportion of mentioned target chosen than either the definite form or the demonstrative. The definite and demonstrative referential forms are not different from each other. All referential forms had significantly lower proportions of mentioned referent choices in the ‘click’ conditions than in either ‘blank’ or ‘look’ conditions. Note also that in the ‘blank’ and ‘look’ conditions there is no effect of the referential form used – the proportions of mentioned referent choices are not significantly different. However, in the ‘click’ condition, the indefinite form elicits a significantly smaller proportion of choices of the mentioned referent than the other two forms. Overall, the general preference of the indefinite form for the mentioned referents was preserved across all conditions, only in the ‘click’ case was the proportion slightly less than 50% – around 43%.

A more detailed examination of the proportion of referent choices for the indefinite referential form condition (Figure 2.6b), comparing mentioned and non-mentioned referents, confirms a preference for the mentioned (discourse active) entity across all intervening material conditions. A repeated measures ANOVA on the transformed data shows main effects of mention, F(1,17)= 31.251, p < 0.001, and the type of intervening material, F(1,17)= 8.876, p = 0.008, as well as a strong interaction, F(1,17)= 16.834, p < 0.001. The mentioned referent choice proportions are significantly greater for the ‘blank’ and the ‘look’ conditions than for ‘click’ condition, and the non-mentioned referent choice proportions are significantly smaller for the same comparison.
Figure 2.6: Referent Choices, Experiment 2. (a) Top Panel: Percent of mentioned referent choices across different intervening material conditions for all three referential forms (that N, the N, a(n) N). (b) Bottom Panel: Percent of choices for mentioned and non-mentioned referents in the indefinite reference form condition. Error bars represent SEM.
Figure 2.7: Fixation Proportions on the Mentioned Referent. Proportions of fixations over time to the mentioned referent across different intervening material conditions (blank, look, click), grouped by the referential form used in the test instruction (that N, the N, a(n) N). Only trials in which participants were not looking at the mentioned referent at the onset of the determiner are included.
Figure 2.8: Fixation Proportions on the Non-Mentioned Referent.
Proportions of fixations over time to the non-mentioned referent across different intervening material conditions (blank, look, click) grouped by the referential form used in the test instruction (that N, the N, a(n) N). Only trials in which participants were not looking at the non-mentioned referent at the onset of the determiner are included.
Eye-movements

The analysis of the initial eye movements during the discourse setup sentence reveals the same fixation patterns as in Experiment 1 (Figure 2.3), confirming a proper discourse setup procedure.

Figures 2.7 and 2.8 plot the proportion of eye fixations on the mentioned referent (Lemon that was the location of the first instruction) and the non-mentioned competitor (the other Lemon in the display) across the three intervening material conditions. The total proportions do not sum to 1.0 because fixations to the other items on the screen are not plotted. The zero point represents the onset of the determiner, with each trial aligned separately based on individual onset measurements. In this experimental setup there is no baseline difference in fixation proportion to the mentioned referent, like in Experiment 1. Due to the test target being a location in the initial discourse setup sentence, participants’ gaze usually moves off of it when the target of the setup sentence has been identified. However, for the purposes of continuity and ease of interpreting the results, we have aligned our data in the same way as we did for Experiment 1; we only include trials in which the eye gaze is not located on the possible referents at the moment of the determiner onset.

In all three referential forms conditions, we see a rise in proportions of fixations to the target within 200 ms, which is consistent with the beginning of signal driven reactions. We have isolated trials for the different form conditions (‘that N’, ‘the N’, ‘a N’) into separate graphs. Figure 2.7 plots the proportions of fixations to
the mentioned location of the initial sentence, and Figure 2.8 – to the non-mentioned location double. A repeated measures ANOVA was performed on the transformed fixation proportions data over three time windows of 200-600 ms, 600-1000 ms and 1000-1400 ms.

For the fixations to the mentioned target (Figure 2.7), the first time window (200-600ms) reveals a significant main effect of referential form, $F(1,17)= 8.670, p = 0.009$, and of intervening discourse type, $F(1,17)= 20.297, p < 0.001$, but no interaction.

Considering the individual referential form conditions, in the ‘that $N$’ condition, we find evidence that in the ‘blank’ condition fixation proportions rise earlier and therefore higher than in the ‘look’ and ‘click’ conditions ($F(1,17)= 24.326, p < 0.001; F(1,17)= 14.545, p = 0.001$), but ‘look’ is not significantly different from ‘click’. In the ‘the $N$’ condition, the results are the same as for the ‘that $N$’. In the ‘blank’ condition, the fixation proportions rise earlier and therefore higher than in the ‘look’ and ‘click’ conditions ($F(1,17)= 45.353, p < 0.001; F(1,17)= 8.920, p = 0.008$), but ‘look’ is not significantly different from ‘click’. Finally, in the ‘a $N$’ case the analysis does not reveal any significant fixation proportions differences among conditions.

The first window (200-600 ms) for the non-mentioned target fixations (Figure 2.7), reveals only a main effect of the intervening material type, $F(1,17)= 5.320, p = 0.034$. Individually, all intervening type conditions are different from each other – ‘blank’ has a greater proportion of fixations than ‘look’, $F(1,17)= 5.396, p =
0.033, ‘click’ has a greater proportion of fixations than ‘blank’, F(1,17)= 5.320, p = 0.034, and ‘click’ has a greater proportion of fixations than ‘look’, F(1,17)= 21.398, p < 0.001.

Analysis of the second time window (600-1000 ms) for the fixations to the mentioned target (Figure 2.7), reveals, same as for the first window, a significant main effect of referential form, F(1,17)= 10.102, p = 0.006, and of intervening discourse type, F(1,17)= 56.646, p < 0.001, but no interaction.

When we examine each referential form, the results are largely similar to those from window one. In the ‘that N’ condition, we find evidence that the ‘blank’ condition has a greater proportion of fixations that the ‘look’ and ‘click’ conditions (F(1,17)= 24.458, p < 0.001; F(1,17)= 29.515, p < 0.001), while ‘look’ is not significantly different from ‘click’. ‘The N’ condition is the same – the ‘blank’ condition has a greater proportion of fixations that the ‘look’ and ‘click’ conditions (F(1,17)= 26.273, p < 0.001; F(1,17)= 13.499, p = 0.002), while ‘look’ is not significantly different from ‘click’. Finally, in the ‘a N’ condition only the ‘blank’ condition is significantly different from the ‘click’ condition, F(1,17)= 7.173, p = 0.016, with the other differences not being significantly different from each other.

An analysis of the second time window (600-1000 ms) for the non-mentioned target fixations (Figure 2.8), reveals significant main effects of referential form, F(1,17)= 15.069, p = 0.001, and of intervening discourse type, F(1,17)= 21.484, p < 0.001, but no significant interaction.
Examining the individual referential form conditions, for ‘that N’ condition, we find that the ‘click’ condition has a significantly greater proportion of fixations than the ‘blank’ and ‘look’ conditions (F(1,17) = 6.857, p = 0.018; F(1,17) = 6.186, p = 0.024). In ‘the N’ condition, only the ‘blank’ condition has significantly less fixation proportions than the ‘look’ and ‘click’ conditions (F(1,17) = 6.124, p = 0.024; F(1,17) = 18.410, p < 0.001). For the ‘a N’ condition, we find no significant differences between conditions due to a delayed rise to target timing.

However, examining the third time window of 1000-1400 ms for the non-mentioned target fixations (Figure 2.8), we find that in ‘that N’ condition all fixation proportions line up on top of each other and no significant differences can be found. ‘The N’ condition is essentially the same as in window two – only the ‘blank’ condition has significantly less fixation proportions than the ‘look’ and ‘click’ conditions (F(1,17) = 6.380, p = 0.022; F(1,17) = 15.201, p = 0.001). For the ‘a N’ condition, we now find a significant advantage of the ‘click’ condition over the other conditions (F(1,17) = 13.012, p = 0.002; F(1,17) = 8.239, p = 0.01).

**Discussion**

The results of Experiment 2 (Figures 2.7 and 2.8) are very similar to those of Experiment 1. There is a clear advantage of the ‘blank’ condition over the ‘look’ and ‘click’ conditions across all intervening material conditions in terms of greater fixation proportions, a faster rise to target and a prevailing preference of all referential forms for mentioned referents.
Overall, the effects of the discourse manipulation on the results for the ‘that N’ and ‘the N’ conditions conformed to our predictions. For ‘that N’ condition, we find that for all types of intervening material, participants had a strong preference in favor of the mentioned location item with little competition from the alternative non-mentioned location item or the non-mentioned cohort competitor. For ‘the N’ form, we also find a strong preference for the mentioned target item. However, there is a significant proportion of fixations to the non-mentioned competitor in the ‘click’ condition. This confirms the trend for the increased intervening discourse to increase competition from non-mentioned alternative referents that was observed in Experiment 1.

The indefinite condition, however, again did not conform to our predictions and is again demonstrating a strong preference for the mentioned referent, which is significantly diminished in the case of a new entity being added into the discourse model (the ‘click’ condition). This preference of the indefinite reference form for a mentioned referent is maintained in the ‘blank’ and ‘look’ conditions with no clear preference in the ‘click’ condition. However, in the ‘click’ condition, the level of fixation proportions to the mentioned referent is still much greater than allowed for by the our predictions – both mentioned and non-mentioned referents are equally considered with both reaching about 40% of fixation proportions.

The lack of a strong referent preference in the ‘click’ condition is due to the salience of the mentioned location being decreased due to the intervening discourse material, which also appears to modulate the timing of the proportions’ rise to target.
There is also more competition from the non-mentioned alternative referent compared to Experiment 1, particularly in the definite and indefinite form conditions which is due to the lesser initial salience of the Location item compared to the Theme of the initial instruction.

Overall, the introduction of a new discourse entity in the ‘click’ condition decreased the accessibility of the mentioned Location referent as manifested in the delayed fixations to the referent and the increased proportion of fixations to the non-mentioned competitors in the definite and indefinite form conditions.

On the basis of the results from Experiments 1 and 2 we can conclude that: (a) discourse changes influence the accessibility of previously mentioned entities; (b) introduction of a new discourse entity leads to a decrease in salience of established entities; (c) different referential forms are affected by the discourse differently: the demonstratives and definites have a more rigid interpretation, while the indefinites are much more susceptible to changes in interpretation; and (d) discourse effects are modulated by the initial salience level of the referent.

The finding of the persisting preference of the indefinite referential form for the mentioned referents is somewhat surprising. Two possible explanations suggest themselves.

The first is that the indefinite form is not interpreted as a non-unique/non-mentioned discourse entity, but rather licenses an existential ‘any’ interpretation. Participants possibly interpreted ‘Click on a NP’ to mean ‘Click on any object (old or new) that satisfies the denotation of the noun’, as in the following example: “Pick
a card!", which means pick any card. In this case the preference for the mentioned items would be due to their higher salience and the ease of locating them in the display. Under this scenario, the choice of the indefinite referential expression would be largely specified by the form-specific licensing conditions and not by the level of the potential referents’ salience.

The second possibility is that ‘a N’ case is indeed interpreted as a non-unique/non-mentioned discourse entity as described by the Givenness Hierarchy, but the setup of these experiments did not fit the preferred felicity conditions for this form’s use. Perhaps, the discourse model must contain at least two entities of similar salience levels in order for ‘a N’ to refer felicitously to a discourse-new entity.

Consider these examples:

Example:

**good** – equal salience
Lisa put two plates on the table. [ two plates on display ]
She then moved a plate to the chair. [ plates are of equal salience ]

**bad** – salience difference
Lisa put a plate next to a jar. [ two plates on display ]
#She then moved a plate to the chair. [ because the mentioned plate is more salient ]

Experiment 3 is designed to discriminate between these alternatives by adding a third exemplar of the type of entity referred to during the discourse setup by the location phrase. This creates a situation where there is both a salient mentioned entity (the lemon referred to) and two non-mentioned alternatives (the other two lemons) that should be of equal salience and less salient than the mentioned entity.
Chapter 3 – Processing of Indefinite Referential Form

Based on the results of the two previous experiments, it seems that in situations when the discourse dynamically changes, different anaphoric forms only partially conform to the predictions of the Givenness Hierarchy (Gundel et al., 1993). On the one hand, as predicted, the preference for referring to a salient mentioned entity is strongest for *that* N, followed by *the* N, with *a(n) N* showing the weakest preference, especially with a new intervening discourse entity. On the other hand, listeners preferentially interpret the indefinite form as referring to a salient discourse entity, which is inconsistent with the Givenness Hierarchy.

We have discussed the two possible explanations, one of which highlights the importance of considering factors specific to each referential form, as argued for by Kaiser (2003), rather than a single global measure of salience. The first possibility is that the indefinite form is interpreted to mean ‘*any* N’, without much consideration of potential referents’ salience status. The second possibility is that the indefinite assumes more than one entity of the same salience level – here the salience is the main influencing factor, but with an additional context requirement. For results consistent with this interpretation, see Chambers, Tanenhaus, Eberhard, Filip and Carlson (2002). In Experiments 1 and 2, we had one mentioned entity, which is by assumption highly salient, and one non-mentioned entity, which is presumably much lower in salience. Thus, the felicity conditions might be violated if the indefinite form requires a choice from among several entities of roughly the same salience –
usually it would be a relatively low salience level, such as for a non-mentioned, discourse-new entity.

To test this hypothesis, we changed our experimental setup so that the felicity conditions would be met. Previously, our displays contained only two identical objects that were potential referents for the test instruction. We now introduce a third identical non-mentioned object into the display. The participants’ now have a choice between a target object mentioned in the setup sentence, and two non-mentioned objects identical to the first one scattered around the display. If the felicity conditions for the indefinite form were being violated before, they should now be satisfied because there are two entities of the same low salience level.

**Experiment 3** – 3 lemons experiment

**Methods**

The experimental setup, the auditory stimuli and the procedure remained exactly the same as in Experiment 2. The test instructions continue to refer to the Location (lemon), not the Theme (heart) as in (1’-3’). The display setup was changed to include a third competitor (lemon): the cohort items were removed and replaced by the third lemon, and a third identical copy of a distracter object pair (Figure 3.1). The complete display now contains: an initial instruction target (Theme) and its pair (two hearts), the initial target’s location (Goal) and its two copies (three lemons), two
other pairs of objects, one other triplet of objects, three unrelated distracters and a cross for the ‘look’ condition.

Figure 3.1: Sample Display, Experiment 3.
Participants

18 male and female University of Rochester undergraduate students who volunteered to participate in the experiment and were paid $7.50. All participants were native speakers of English, had normal or corrected to normal vision and no reported hearing difficulties.

Predictions

If the reason for the absence of a preference for the low salience entity in Experiments 1 and 2 was that there was only one low salience alternative, then our predictions are the same as they were for Experiment 2. The demonstrative and definite forms should prefer the mentioned referent to the non-mentioned ones resulting in a higher proportion of choices and fixations on the mentioned Location item. The indefinite form should prefer the non-mentioned items as a referent with a higher probability of choosing one of them as the test instruction target and a greater proportion of fixations on both non-mentioned competitors than on the mentioned one. As before, we expect that an intervening sentence that introduces another entity into the discourse will result in a diminished salience of existing discourse entities and a greater flexibility in the choice of the referent by different referential forms with the demonstrative form not being affected much, while the indefinite form should change its interpretation more significantly.
Figure 3.2: Referent Choices, Experiment 3. Top panel: Percent of mentioned referent choices across different intervening material conditions for all three referential forms (that N, the N, a(n) N). Bottom Panel: Percent of choices for mentioned and non-mentioned referents in the indefinite reference form condition. Error bars represent SEM.
Figure 3.3: Fixation Proportions on the Mentioned Referent. Proportions of fixations over time to the mentioned referent across different intervening material conditions (blank, look, click), grouped by the referential form used in the test instruction (that N, the N, a(n) N). Only trials in which participants were not looking at the mentioned referent at the onset of the determiner are included.
Figure 3.4: Fixation Proportions on the Non-Mentioned Referents. Proportions of fixations over time to the non-mentioned referent across different intervening material conditions (blank, look, click) grouped by the referential form used in the test instruction (that N, the N, a(n) N). Only trials in which participants were not looking at the non-mentioned referent at the onset of the determiner are included.
Results

Choice of referent

Figure 3.2a presents the proportion of choices of the mentioned location when it was selected as the referent of the test sentence. It displays proportions for the different referential conditions (that N, the N and a N) across the three intervening material conditions (‘blank’, ‘look’ and ‘click’). A repeated measures ANOVA was performed on the transformed data, as in the previous experiments.

The overall pattern of results in this experiment is similar to Experiment 2. The analysis revealed significant main effects of referential form, F(1,17)= 70.351, p < 0.001, and intervening material type, F(1,17)= 36.725, p < 0.001, but no significant interaction. The proportion of the mentioned entity choices is significantly greater in the ‘blank’ condition compared with the ‘look’ and ‘click’ conditions for all anaphoric form types, with marginal significance for ‘that N’ condition: ‘a N’ ‘blank’ vs ‘look’, F(1,17)= 6.174, p = 0.024, ‘a N’ ‘blank’ vs ‘click’, F(1,17)= 17.113, p < 0.001; ‘the N’ ‘blank’ vs ‘look’, F(1,17)= 6.127, p = 0.024, ‘the N’ ‘blank’ vs ‘click’, F(1,17)= 20.718, p < 0.001; ‘that N’ ‘blank’ vs ‘look’, F(1,17)= 3.754, p = 0.069, ‘that N’ ‘blank’ vs ‘click’, F(1,17)= 4.341, p = 0.053. Also, the proportion of mentioned referent choices decreases from the demonstrative form to the definite form (marginal significance), F(1,17)= 3.782, p = 0.069, and from the definite to the indefinite form, F(1,17)= 20.586, p < 0.001. This trend is modulated by the intervening material, with the largest difference among the three referential forms in the ‘click’ condition. Overall, we continue to observe the general preference
for the mentioned referent across all reference forms (proportion of mentioned referent choices > 50%).

A more detailed examination of the proportion of referent choices for the indefinite referential form condition (Figure 3.2b), comparing the mentioned entity with the non-mentioned identical referents, confirms a preference for the mentioned (discourse active) entity across all intervening material conditions. A repeated measures ANOVA on the transformed data demonstrates a main effect of mention, F(1,17)= 35.004, p < 0.001, no effect of the type of intervening material, and an interaction, F(1,17)= 14.388, p = 0.002. The mentioned referent choice proportion is significantly greater in the ‘blank’ condition than in ‘look’ or ‘click’ conditions (F(1,17)= 6.174, p = 0.024; F(1,17)= 17.113, p < 0.001) which is different from Experiment 2 where conditions ‘blank’ and ‘look’ differed from ‘click’. The trend in the non-mentioned referent choices was also different from the previous experiment. In Experiment 2 the proportion of non-mentioned referent choice was greater in the ‘click’ condition, whereas now this proportion is not significantly different across all intervening material types. For the demonstrative form, the sum of the proportions of non-mentioned choices results in about 25% of choice in favor of the non-mentioned referent vs 75% of choice preferring the mentioned one. For the definite form, we obtain 42% for non-mentioned vs 51% for the mentioned one. For the indefinite form, the sum of the two non-mentioned entities is similar to the mentioned referent – 44% vs 42%.
Eye-movements

Figures 3.3 & 3.4 plot the proportion of fixations on the mentioned referent (e.g., the lemon that was the Location of the first instruction) and the non-mentioned competitors (e.g., the two other lemons in the display) across the three intervening material conditions. The total proportions do not sum to 1.0 because the fixations to other items on the screen are not plotted. The zero point represents the onset of the determiner, with each trial aligned separately based on individual onset measurements. As in Experiment 2, there is no baseline difference in fixation proportions to the mentioned referent at the onset of the test instruction. However, for the purposes of continuity and ease of interpreting the results, we have aligned our data at the determiner and are plotting only trials in which the eye gaze is not located on any of the three possible referents at the moment of the determiner onset.

In all three referential forms conditions, there is a rise in fixation proportions to the target within 200ms of the determiner onset, which is consistent with when we first expect to see signal-driven fixations. We have isolated trials for the different conditions (‘that N’, ‘the N’, ‘a N’) into separate graphs. Figure 3.3 plots proportions of fixations to the mentioned location of the initial sentence, and Figure 3.4 – to the sum of the two non-mentioned alternatives, providing us with the total proportions of fixations on non-mentioned referents. As in the previous experiments, a repeated measures ANOVA was performed on the transformed fixation proportions data over two time windows: 200-600ms and 600-1000ms. The window length is the same as the average noun duration in the test sentence.
For the fixations to the mentioned target (Figure 3.3), the first time window (200-600ms) reveals significant main effects of referential form, $F(1,17)= 11.342$, $p = 0.004$, and of intervening material type, $F(1,17)= 79.884$, $p < 0.001$, but no interaction.

Considering the individual referential form conditions, we find a significant difference mainly between ‘blank’ and the other two intervening material conditions for all reference forms. The fixation proportions to the mentioned referent in the ‘blank’ condition rise faster and are greater: ‘that $N$’ condition – ‘blank’ vs ‘look’, $F(1,17)= 29.681$, $p < 0.001$; ‘blank’ vs ‘click, $F(1,17)= 18.768$, $p < 0.001$, ‘the $N$’ condition – ‘blank’ vs ‘look’, $F(1,17)= 34.895$, $p < 0.001$; ‘look vs ‘click, $F(1,17)= 12.433$, $p = 0.003$, ‘a $N$’ condition – ‘blank’ vs ‘look’, $F(1,17)= 31.894$, $p < 0.001$; ‘blank’ vs ‘click, $F(1,17)= 31.077$, $p < 0.001$.

The first window (200-600ms) for the non-mentioned target fixations (Figure 3.4), reveals only a main effect of the intervening material type, $F(1,17)= 10.002$, $p = 0.006$. Individually, only the proportion of fixations to the non-mentioned referent in ‘that $N$’ condition is greater in the ‘click’ condition than in the ‘blank’, $F(1,17)= 7.776$, $p = 0.013$.

An analysis of the second time window (600-1000ms) for the fixations to the mentioned target (Figure 3.3), also reveals a significant main effect of referential form, $F(1,17)= 13.024$, $p = 0.002$, and of intervening discourse type, $F(1,17)= 96.656$, $p < 0.001$, but no interaction.
On the level of individual referential form conditions, the statistical patterns are the same as those from the first window. For all referential forms, the fixation proportions to the mentioned referent in the ‘blank’ condition are greater than for the ‘look’ or ‘click’ conditions, which are not different from each other: ‘that N’ condition – ‘blank’ vs ‘look’, F(1,17)= 47.098, p < 0.001; ‘blank’ vs ‘click, F(1,17)= 41.973, p < 0.001, ‘the N’ condition – ‘blank’ vs ‘look’, F(1,17)= 19.800, p < 0.001; ‘look vs ‘click, F(1,17)= 30.452, p < 0.001, ‘a N’ condition – ‘blank’ vs ‘look’, F(1,17)= 21.525, p < 0.001; ‘blank’ vs ‘click, F(1,17)= 40.755, p < 0.001.

For the second time window (600-1000ms) fixations on the non-mentioned referents (Figure 3.4), are similar to those for the mentioned referent fixations (Figure 3.3). There are significant main effects of referential form, F(1,17)= 6.967, p = 0.017, and of intervening discourse type, F(1,17)= 51.545, p < 0.001, but no interaction.

We now consider each referential form. The fixation proportions to the non-mentioned referents in the ‘blank’ condition are much greater than for ‘look’ and ‘click’, which are not different from each other: ‘that N’ condition – ‘blank’ vs ‘look’, F(1,17)= 11.124, p = 0.004; ‘blank’ vs ‘click, F(1,17)= 15.665, p = 0.001, ‘the N’ condition – ‘blank’ vs ‘look’, F(1,17)= 9.995, p = 0.006; ‘look vs ‘click, F(1,17)= 20.200, p < 0.001, ‘a N’ condition – ‘blank’ vs ‘look’, F(1,17)= 7.851, p = 0.012; ‘blank’ vs ‘click, F(1,17)= 24.405, p < 0.001.
**Discussion**

The results of Experiment 3 (Figures 3.2 – 3.4) indicate that the preference for the mentioned entity in the indefinite condition was not due to there being only one low salient alternative. Even when there were two low salient alternatives, listeners did not prefer the low salience entity to the high salience entity. Thus, it appears that the indefinite form was interpreted to mean ‘*any N*’ that satisfies the referential description. This result strongly suggests that, for the indefinite form, salience of the potential referents does not play a significant role in determining the appropriate form of the reference, but rather specific constraints pertinent to the form affect that choice.

The results for the ‘*that N*’ and ‘*the N*’ conditions conformed to our predictions. For ‘*that N*’ condition, we find that for all types of intervening material, participants had a strong preference for the mentioned location item as the referent with little competition from the alternative non-mentioned referents. While this behavior is consistent with the Givenness Hierarchy predictions, on the basis of Experiments 1-3 it appears that the demonstrative form may have a strong requirement for a mentioned referent as its preferences did not change depending on the initial salience level of the discourse entity.

For ‘*the N*’ form, we also find a strong preference for the mentioned referent item. However, this time there is a significant proportion of fixations to the non-mentioned competitors in the ‘look’ and ‘click’ conditions. This confirms the trend
for the increased intervening discourse to increase competition from non-mentioned alternative referents which was observed in previous experiments. This form, based on our results, does not appear to have any strong non-salience-based constraints on its interpretation or use.

The indefinite condition ‘a N’, is again demonstrating a preference for the mentioned referent, which is again significantly diminished in the case of another entity being added into the discourse model (the ‘click’ condition). This overall preference of the indefinite referential form for the mentioned referent is maintained in the ‘blank’ condition, with no clear preference for mentioned or non-mentioned referents in the ‘look’ and ‘click’ conditions, contrasting with Experiment 2. This lack of a strong referent preference in the ‘look’ and ‘click’ conditions is likely due to the fact that there are two non-mentioned items in the display and only one mentioned item.

Overall, there are three main results from this experiment:

1) The lower salience of the target discourse entity (Lemon) due to the presence of multiple non-mentioned competitors compared with Experiment 2, affected the intermediate condition ‘look’ to a greater degree: in Experiment 2, ‘look’ condition patterned with the ‘blank’ condition, whereas now it patterns with the ‘click’ condition. This suggests that the initial lower salience allows for an increased discourse interference – the mentioned referent is being resolved similar to a case when a new entity is introduced to the discourse rather than similarly to the no intervening material case.
2) There is a strong trend for the proportion of mentioned referent choices to decrease with the increase in the intervening material: not significantly for the ‘that N’ condition, greater for the ‘the N’, and the most significantly for the ‘a N’ condition. Overall, the introduction of a new discourse entity in the ‘click’ condition decreased the accessibility of the mentioned Location referent as manifested in the delayed fixations to the referent and the increased proportion of fixations to the non-mentioned competitors in the definite and indefinite form conditions.

3) Finally, while we find strong influence of discourse salience on some aspects of anaphor resolution, we also find strong evidence that form-based constraints take precedence in certain situations for some referential forms.

In sum, our results demonstrate that while salience is an important factor affecting the preference for different referential expressions, and the overall ordering is consistent with the Givenness Hierarchy, it is not the only factor that determines the interpretation and use of a referring expression. Rather, one must also consider specific properties of the referential expression itself, as argued for by Kaiser and colleagues (2008). Clearly the indefinite form can be interpreted as referring to any matching entity, without much regard to its salience. It also seems likely that there is a particular interpretation associated with some of the other referential forms we used. In particular, ‘that N’ might have a contrastive interpretation. We address this in passing when discussing the results of the next experiment.
Chapter 4 – Reference Processing in a Narrative Discourse

In previous experiments we have demonstrated that different anaphors react differently to dynamically changing discourse conditions. As predicted, the introduction of new discourse relevant material between the initial mention and the subsequent reference changed the preferred referent differentially for different anaphoric forms – the most for the indefinite form, and the least for the demonstrative. These results suggest that there are different constraints of varying strength tied to individual referential forms that affect the process of referent choice in addition to the referent’s mental salience. Also, all of the forms preferred the mentioned referent, which is only partially consistent with the predictions of the Givenness Hierarchy (Gundel et al., 1993).

Most of these form-specific properties and licensing conditions are linked to speaker intentionality or situational context which means that the circumstances and the task in which a referential form is used should have a strong impact on its interpretation. In the setup for this experiment we aim to examine how a task change from an explicit referent choice to a scene verification judgement would affect the effects and interpretations we found in prior studies. We continue to investigate the behavior of the selected referential forms (that N, the N, a(n) N) under different developing discourse conditions, but this time in a more natural linguistic setting and using more structured visual displays. Finally, we also include manipulations
intended to evaluate the relative salience of the referents of referring expressions
directly through non-linguistic means.

In Experiments 1-3 participants viewed a display that was an object array
with a large number of items in it, which carries certain disadvantages with it. It is a
cluttered visual environment that hinders visual search for referents, and it is an
artificial scene representation which restricts contextual interpretation of the
reference. The displays we utilize here allow us to partially remedy these concerns
by creating a structured environment to facilitate visual search – participants view
schematic scenes of rooms with furniture pieces and several objects randomly
distributed on the furniture and on the floor (see Figure 4.1). The linguistic input is
now a natural narrative style discourse without breaks or pauses which describes the
scene in terms of how a person interacts with it and what objects that person moves
and where. Following each sentence, an object explicitly moves within the displayed
scene. The participants’ task is to verify whether the scene change matches the
description they heard. Introduction of these dynamic changes allows us to
manipulate the alternative referents participants have to consider to make their
response.

Overall, we have two main goals for this experiment. (1) One is to assess the
effects a task change would have on referent preferences – we compare an explicit
referent choice with a judgement of what referent is acceptable to different anaphors.
The first task is more referential in that it forces a specific single and unique
interpretation of the reference, whereas the second task is more loosely defined and
should highlight the relative strength of the goodness of fit of alternative referents to individual referential expressions. (2) We also aim to examine whether the three referential forms we use have significantly different licensing conditions on their use and interpretation as previous studies suggest. Using dynamic scene changes should allow us to examine the goodness of fit to the different anaphors of potential referents of high and low discourse salience independently of each other rather than pitching them in direct competition.
Sample Discourse:
0) The kitchen was a mess, so Lisa decided to clean up.
1) She moved a bowl to the table.
2) She then placed the cactus in the cupboard.
3) Then, she put {a / the / that} bowl in the cupboard.

**Figure 4.1: Sample Display, Experiment 4.** Solid lines denote movements of mentioned objects. The dotted line denotes a possible movement of a non-mentioned competitor object during the test sentence. Three intervening material conditions are possible between the setup and the test sentences: (a) 'blank' condition - when nothing is intervening and the first sentence is followed directly by the last one (indicated by arrow), (b) 'flash' condition - when there is no intervening sentence, but a grey circle 50 pixels in diameter briefly flashes for 50ms in some location on the screen, (c) 'move' condition - when an intervening sentence describes a move of an object which is then followed by the actual move on the screen.
**Experiment 4**

In this experiment we manipulate the discourse change in the same way as before – through introduction of various degrees of intervening material between the initial discourse setup and the test sentence. We also keep the same three referential forms (*that* N, *the* N or *a(n)* N), with the test sentence always referring to the Theme of the setup sentence. We assume that with more intervening material the initial mental representation of the Theme will become less salient, and consequently the entity it indicates should be less accessible to the referring expression, leading to it being less acceptable as the intended referent.

In Experiments 1-3 participants clicked with the computer mouse on the referent they considered the best match to the instruction they heard. In this experiment we use a scene verification task in which participants listen to a description of a scene, that details the changes happening with the scene, and then make a judgement as to whether the description matches the scene. The scene description is a narrative of several sentences that describe actions some person (Lisa) performs with the objects in the scene. Following each sentence, an explicit change is displayed in the scene: an object moves to a new location, each move is explicit – an object jumps to its new location inducing a kind of apparent motion effect, and the move only occurs after the offset of the auditory sentence. The following is how the task was described to the participants:
“You will hear a description of actions that Lisa performed. On the screen, the appropriate changes will happen to the scene. However, on some trials, the change on the screen will not match the description of Lisa’s action that you hear. Your task is to push a Yes or No button to indicate at the end of each sentence whether the change matched the description. The experiment does not pause while you respond, so you should do it promptly.”

The narrative consisted of three or four sentences describing consecutive actions by a person (Lisa) interacting with various, sometimes repeating, objects in the scene. (0) The first sentence is used to situate the discourse in the scene – it mentions the type of room Lisa is in (kitchen, office, living room, recreation room, bedroom, dining room), and what she is doing – cleaning, organizing or rearranging things, on half of the trials also giving a reason for her actions. For example, “The kitchen was a mess, so Lisa decided to clean up” or “Lisa was organizing the office”. (1) The next sentence, the discourse setup sentence, introduces the antecedent and is always of the form, “She moved a(n) X(object) to the Y(furniture piece)”. (2) The intervening discourse sentence manipulates the change in salience of the discourse entities from the setup sentence by introducing a new discourse entity between the antecedent and the anaphor. (3) The test sentence always comes last and refers to the Theme entity mentioned in the setup sentence (1), using one of the three anaphoric forms: that N, the N or an N. A sample discourse is presented in Table 4.1 (0)–(3) for the sample display in Figure 4.1.
Table 4.1: Sample Discourse, Experiment 4

(0) [lead-in sentence] The kitchen was a mess, so Lisa decided to clean up.

(1) [discourse setup sentence] She moved a bowl to the table.

(2a) [intervening material] <blank>
(2b) <flash>
(2c) She then placed the cactus in the cupboard.

(3a) [test sentence] Then, she put that bowl in the cupboard.
(3b) Then, she put the bowl in the cupboard.
(3c) Then, she put a bowl in the cupboard.

To manipulate the salience of the initial discourse representation, we used one of three types of intervening material (2)(a-c) between the setup sentence (1) and the test sentence (3), similarly to previous experiments:

(2a) ‘Blank’ condition – no intervening sentence, the final test sentence follows immediately after the first one. This does not change the initial salience distribution in any way and serves as a baseline.

(2b) ‘Flash’ condition – a brief attention-capture manipulation to divert the participants’ eye gaze from the mentioned items without introducing any new linguistic material and without a break from the narrative.

(2c) ‘Move’ condition – an intervening sentence that introduces a new discourse entity while maintaining coherence of the discourse and the task – Lisa performs another action in the scene (moves an object). This condition changes the mental status and salience representation of the entities within the discourse model.
In the ‘flash’ condition, we capture participants’ eye gaze and attention to a specific location within the scene by a sudden-onset, briefly flashed spatial cue (Jonides & Yantis, 1988). This sub-threshold flash occurs on the placement surface that the intervening sentence (2c-\textit{move}) would have referred to. This results in a “look-away” condition equivalent in eye position to the intervening discourse condition (‘\textit{move}’), but without the new linguistic material. The flash itself consists of a small grey circle, subtending an area of approximately 1x1 degree of visual angle, against a white background, onscreen for 50–60 ms (see Figure 4.1 condition b) for illustration). Although some participants reported noticing this cue, they assumed it was a technical problem and not a part of the experiment. It proved quite effective in capturing attention – a subsequent analysis showed a significant increase in the proportion of fixations to the location of the flash immediately following its onset.

In sentence (1), both bowls (on the floor and on the counter) are possible referents introduced by the sentence. However, following the display change showing the move by the bowl on the floor, this item becomes the Theme of the setup sentence and the higher-salience one of the two bowls. This manipulation provides us with two identical objects of different discourse salience which we can investigate by directing attention to them following the test sentence (3) by using scene changes – either one of these bowls can move following a test reference.

Some scene changes matched the auditory description, while others did not, with various degrees of mismatch. Overall, there are four possible types of changes,
one matching the description and three not matching: (1) the mentioned item moves to the correct location; (2) a non-mentioned item which is the double of the mentioned item moves to the correct location; (3) the mentioned item moves to a wrong location; and (4) a wrong item moves to the correct location. Only the changes (1) and (2) occur in the experimental trials and only following the offset of the final test sentence.

**Predictions**

We expect to confirm the results of previous experiments. The manipulation with the initial salience of the two bowls in the scene gives us two potential referents of varying salience levels – mentioned and non-mentioned. We predict that the mentioned item will be a better referent for all referential forms resulting in a higher proportion of fixations on it and a high proportion of match responses to mentioned item move scene changes. When the non-mentioned item moves following the test sentence, we predict a differential effect for the referential forms used – with the indefinite form, participants should accept both mentioned and non-mentioned referents as a good match to the scene description, the demonstrative should strongly prefer the mentioned referent, and the definite form should fall somewhere in between.

The response times should reflect the difficulty of the participants’ choice and should generally be slower to accept a non-mentioned item as a referent, and should differ across the referential forms. The better the fit between a referent and
the referential expression, the faster the response times. For the changes involving a mentioned item, the response times should be similar across conditions, while for the changes involving a non-mentioned item, the indefinite form should have roughly the same response time for the mentioned and non-mentioned referents, while the demonstrative and the definite forms should have longer reaction times when accepting the non-mentioned items as the referent. However, when rejecting a moved item as a referent – the mismatch condition – the reaction times should generally be shorter across all referential forms.

As in the previous studies we expect that an intervening sentence that introduces a new entity into the discourse will result in a diminished salience of the existing discourse entities and a greater flexibility in the choice of the referent by different referential forms, with the demonstrative form not being affected much, while the indefinite form should change its interpretation more significantly.

**Methods**

**Participants**

Twenty male and female University of Rochester undergraduate students who volunteered to participate in the experiment were paid $10 for their time. All participants were native speakers of English, had normal or corrected to normal vision and no reported hearing difficulties.
Materials

All auditory stimuli were pre-recorded in a noise-proof laboratory on a DAT tape and then digitized at 44kHz. All trials were recorded individually with all sentences following after each other to avoid any discontinuity of pitch, voice gaps or irregularities. The speaker was instructed to use normal intonation and avoid placing emphasis on any part of any sentence to avoid biased interpretations. The first mention of any paired object in the scene is always with an indefinite article. The Theme of the discourse setup sentence is always located on the floor – Lisa moves it to one of the furniture pieces in the room. In experimental trials, each sentence mentions a different placement area, thereby covering the whole display. The final test sentence on experimental trials always refers to the Theme of the initial sentence using varying referential forms. There are three verbs that are randomly used to describe Lisa’s actions – move, place and put.

The visual display (Figure 4.1) represents a schematic scene of a room with several objects scattered around and three different pieces of furniture (e.g. table, counter, cupboard). There are overall six different scenes/rooms in the experiment that differ from each other in terms of what furniture appears in the scene: kitchen, office, living room, recreation room, bedroom, dining room. The furniture acts as placement surfaces for placement of objects that correspond to Lisa’s actions. Each placement surface is equally likely to occur in sentences (1) through (3). In the initial display, objects can be located anywhere on the three surfaces or on the floor.
between them. However, during the task, objects can only be placed on surfaces, nothing is moved to the floor. Each furniture surface has placement areas for three objects on it, and there are three further placement areas on the floor. However, the placement areas do not form any kind of a grid or a pattern and are different for each of the six scenes. The scenes are centered on the computer screen, with at least 50 pixels of white space around the edges. The display resolution used was 1280 x 1024 pixels. Each placement area was 120x120 pixels in size and subtended approximately 3 degrees of visual angle horizontally and vertically, with at least 20 pixels between the edges of adjacent areas.

Each display contains a total of six items. These items include the item mentioned in the discourse setup sentence (bowl) and its unmentioned identical pair, but of a different color. There is also a distracter object pair (rolling pin) and two other distracters (cactus, wine glass) unrelated to the rest of the items. The displays are structured so that items from the same pair do not appear together on one placement surface. At the beginning of a trial, all areas on the floor are occupied by items (a distracter, and one item from each pair) and each placement surface contains one of the other objects at random. All items are randomly colored line drawings of familiar objects. All objects were 120 pixels on the longest dimension and subtended approximately 3 degrees of visual angle in diameter at a 70 cm viewing distance.

All objects were selected from our special-purpose database so that their names were not phonologically or visually similar, semantically related or associated with each other. All items are reused two to three times during the course of the
experiment – once in each of the roles: the mentioned target of the initial sentence, the distracter object pair and an unrelated distracter. If used in one of the distracter roles, the item can be mentioned during the intervening sentence or in the last test sentence in the filler trials.

The item mentioned in the intervening material sentences (one of the distracters) and the flash circle were positioned at a roughly equidistant location from the target of the discourse setup sentence to equate the visual distance the eye travels when the participant looks away from the discourse Theme to the extent that the setup of the display allows. The placement was always on a different furniture surface than the discourse Theme, and placement surfaces (furniture pieces) were positioned roughly equidistant from each other.

Experimental trials can be of two types: mentioned move trials, where the mentioned object moves to the correct location, and non-mentioned move trials, where the non-mentioned object which is a pair of the mentioned item moves to the correct location.

The experimental trials do not differ from filler trials in terms of either the distribution of objects on the screen or the form of the narrative. However, to disguise the experimental manipulations, filler trials could have errors in any sentence, not just in the last one, although there was still only one error per trial, and a variety of different error types were used. The extra error types for filler trials included movement by a wrong object to a wrong location and a movement by a mentioned object to the wrong location. An error could occur only in one sentence
per trial, but the participants are not made aware of this. Also, within filler trials, the
target of the discourse setup sentence could be either a paired or a unique distracter
item. On some filler trials the final test sentence referred to a referent different from
the Theme of the setup sentence.

All trials were randomized so that no more than two consecutive trials had
either the same intervening material type, the same referential form, the same
location of the initial sentence theme, the same item used as the theme, or the same
type of experimental or filler condition. There were a total of 216 trials with 90
experimental trials (45 correct and 45 wrong), which resulted in five data points for
each of the eighteen conditions: three intervening material conditions x three
referential forms x two types of experimental trials (mentioned item moved and non-
mentioned item moved).

Procedure

An SRI EyeLink II head-band mounted eye-tracker was used to monitor
participants’ eye movements, sampling at a rate of 250Hz. Whenever possible,
attempts were made to track both the pupil and the corneal reflection of both eyes to
improve accuracy. The calibration procedure used a 9-point calibration grid, and a
drift correction was performed before every trial. Experimental stimuli were
presented using custom software written in Visual C++. Pre-recorded auditory
stimuli were delivered through the Sennheiser HD 570s headphones.
The participants’ performed a scene verification task. They responded Yes or No with a button press on the keyboard depending on whether the changes in the displayed scene matched the description of these changes described in the narrative. All participants were comfortably seated approximately 70cm from the computer screen with the monitor at eye level. Each object subtended approximately 3 degrees of visual angle.

Each trial began with the centrally positioned cross on the screen which the participants were required to fixate and then click the computer spacebar to perform the drift correction. 100 ms later the display with the pictures appeared followed by the onset of the auditory stimuli (the lead-in sentence) 200 ms later. The change in the scene always happened immediately following the sentence offset – about 28 ms later. In the ‘flash’ condition, the flash comes on 200 ms following the previous response and stays on for 50–60 ms. The program did not pause to wait for participants’ responses, however there was a short delay of 300 ms between the sentences which was sufficient to provide a response. Following the last (test) sentence the program would wait until a response was provided before advancing to the next trial. The experiment lasted approximately 1 hour excluding the initial explanations, setup and calibration of the eye-tracker.

**Data Collection and Analysis**

The data were collected within the eye-tracker-generated EDF data files, error corrected and then processed using custom made data-analysis software written
in Microsoft Access. Further analysis involved determining the proportion of fixations to each object in 4ms time slices. An eye movement was considered to have landed on an object if the participant fixated anywhere within the item’s placement area.

The eye-movement data provide incremental on-line information about the reference resolution process, while the participant’s responses (matching the scene with its description) indicate the goodness of fit of the referent to the referential form used to refer to it. The reaction times provided a third perspective, indicating the difficulty of making a response judgement. Only trials that did not timeout and did not have any errors were included in the analyses (less than 1% of data was excluded). Trials on which a mentioned object moved were analyzed separately from the non-mentioned object move condition.

Post-experimental debriefing of participants established that none became aware of the experimental manipulations, and thus all participants, who have successfully completed the experiment, were included in the final analysis.
**Figure 4.2: Percent of ‘Match’ Responses to Mentioned Changes.** Proportion of responses indicating a match between the scene change and Lisa’s actions in the condition when the change involved a mentioned referent moving.

**Figure 4.3: Percent of ‘Match’ Responses to Non-Mentioned Changes.** Proportion of responses indicating a match between the scene change and Lisa’s actions in the condition when the change involved a non-mentioned referent moving.
Figure 4.4: RTs of ‘Match’ Responses to Mentioned Changes. Reaction times for trials in which participants gave a ‘match’ response when the change involved a mentioned referent.

Figure 4.5: RTs of ‘Match’ Responses to Non-Mentioned Changes. Reaction times for trials in which participants gave a ‘match’ response when the change involved a non-mentioned referent.
Figure 4.6: RTs of ‘No Match’ Responses to Non-Mentioned Changes. Reaction times for trials in which participants gave a ‘no match’ response when the change involved a non-mentioned referent.
Results

Choice of referent

The choice of referent in this experimental setup is represented by the goodness of fit judgements – whether the change displayed in the scene ‘matches’ the narrative description that was provided. There are two different types of experimental trials, which we are analyzing separately. First, there are trials where the mentioned referent object moved – these always match the discourse description. Second, there are trials during which a non-mentioned object moves. These trials provide the possibility of a mismatch between the listener’s interpretation of the description and the scene. Recall that our previous experiments demonstrated that the definite and the demonstrative forms do not readily resolve to a non-mentioned entity.

Figure 4.2 represents the proportion of ‘match’ responses when the mentioned referent moved across the different referential conditions (that N, the N and a(n) N) by the three intervening material conditions (‘blank’, ‘flash’ and ‘move’). A repeated measures ANOVA was performed on the quasi-empirical logit transformed data, as in the previous experiments. The analysis did not reveal any significant main effects or an interaction. The proportion of positive match responses is close to ceiling with a mean proportion of 97% for that N, 99% for the N, and 93% for a N. Participants consistently judged the scene and its description to match across all intervening material conditions when the test sentence referred to a previously mentioned referent.
The analysis of the response times for the trials in which the change involved a mentioned referent (Figure 4.4) and the participants indicated a ‘match’, demonstrates no significant differences between the discourse manipulation conditions, and a small increase in the response times for the indefinite condition (mean of 799 ms) as compared with the definite form (mean of 700 ms) and the demonstrative form (mean of 695 ms). A repeated measures ANOVA analysis confirms the significance of this difference: $a$ N vs $the$ N, $F(1,19)= 6.065, p = 0.025$; $a$ N vs $that$ N, $F(1,19)= 5.539, p = 0.031$.

Considering that the proportion of data where the participants indicated a ‘mismatch’ is negligible for the mentioned conditions (Figure 4.2), we did not analyze the response times for the ‘no match’ trials.

Figure 4.3 presents the proportion of ‘match’ responses for the conditions in which a non-mentioned item moved. Here, the pattern of results differs from the pattern in the mentioned item move condition. A repeated measures ANOVA performed on the transformed data revealed a main effect of anaphoric form, $F(1,19)= 534.720, p < 0.001$, but no effect of the type of intervening material and no interaction. While the intervening material conditions are not different from each other, the referential form conditions are all significantly different from each other – $that$ N vs $the$ N, $F(1,19)= 12.654, p = 0.002$; $the$ N vs $a$ N, $F(1,19)= 56.802, p < 0.001$; $that$ N vs $a$ N, $F(1,19)= 534.720, p < 0.001$. The indefinite ($a$ N) condition readily resolves to the non-mentioned referent moves (97% match judgments), just as it resolved to the mentioned referent, the definite ($the$ N) form generally rejects
the non-mentioned entity as a referent (only 39% match judgments), and the demonstrative *(that N)* condition strongly rejects the non-mentioned entity – only accepting it as a referent on about 10% of the trials. This strong difference between anaphoric forms is the same across the discourse manipulation conditions.

An analysis of the response times for the trials when the change involved a non-mentioned referent and the participants indicated a ‘match’ (Figure 4.5), also demonstrates no significant differences between the discourse manipulation conditions. This analysis, however, is complicated by the fact that the data is very sparse for the definite and the demonstrative forms – only 39% and 10% of the total trials respectively – which means these data patterns do not reach significance. However, a numeric trend suggests a large increase in the response times (around 300 ms) for the definite form (mean of 1011 ms) and the demonstrative form (mean of 960 ms) as compared with the response times for trials with a mentioned referent change. The indefinite condition (mean of 736 ms) demonstrates a slightly faster response time (by about 60 ms) on the same comparison.

The analysis of the response times for the trials when the change involved a non-mentioned referent and the participants indicated a ‘mismatch’ (Figure 4.6), is also challenged by very sparse data for the indefinite condition (only 3% of data). Examination of the definite and the demonstrative conditions reveals no significant differences between discourse manipulation conditions, and a slight decrease of about 100 ms in response times *(the N – mean of 926 ms, that N – 864 ms)*
compared with responses when participants indicated a ‘match’ for changes involving non-mentioned items (Figure 4.5).

**Eye-movements**

Figures 4.7 and 4.8 plot the proportion of fixations on the mentioned referent (the bowl moved in the discourse setup sentence) and the non-mentioned competitor (the other bowl in the display) across all conditions. Figure 4.7 displays data for the condition when the mentioned referent moves following the test sentence, and Figure 4.8 for when the non-mentioned object moves. The total proportions do not sum to 1.0 because the fixations to other items on the screen are not plotted. The zero point represents the onset of the determiner in the final test sentence, with each trial aligned separately based on individual onset measurements. We are including only those trials in which the eye gaze is not located on any of the possible referents at the moment of the determiner onset.

All of these proportion graphs display similar patterns of eye movements. The preferred referent is fixated shortly after the onset of the target word, and the participants continue to fixate on it until the end of the final sentence when a scene change occurs at approximately 1600 ms (average distance between the determiner and the change onset is 1636 ms). Following the change onset, the participants’ eye gaze is captured by the moving object. This results in a different behavior for the two different change types. When a mentioned referent is moved, (we know from previous experiments that the anaphoric forms we use exhibit an overall preference
for salient mentioned referents), it receives an additional boost in terms of increased fixation proportions. If the non-mentioned referent moves, we observe a shift of fixations from the mentioned item pre-change to the non-mentioned item that moved. Although the effects are smaller than in Experiments 1-3, intervening material increases the proportion of fixations on non-mentioned items with this effect modulated by the anaphoric form. The indefinites have the most fixations and the demonstratives have the least.
Figure 4.7: Fixation Proportions during Mentioned Change Trials. Proportion of fixations over time to mentioned and non-mentioned referents during trials in which the scene change involved the mentioned referent; grouped by referential form (*that* N, *the* N, *a(n)* N) and type of intervening material (blank, flash, move). Fixation proportions are aligned at the determiner onset. Scene change onset occurs at approximately 1600 ms.
Figure 4.8: Fixation Proportions during Non-Mentioned Change Trials.
Proportion of fixations over time to mentioned and non-mentioned referents during trials in which the scene change involved the non-mentioned referent; grouped by referential form (that N, the N, a(n) N) and type of intervening material (blank, flash, move). Fixation proportions are aligned at the determiner onset. Scene change onset occurs at approximately 1600 ms.
**Discussion**

Overall, the results confirm our experimental predictions. The mentioned item was uniformly accepted as the intended referent for all anaphoric forms when it was the target of the last change. In the condition when a non-mentioned item moved in the scene, the indefinite form easily accepted it as the referent, the definite form mostly rejected it (39% match responses), and the demonstrative form strongly rejected it as a potential referent with ‘match’ responses at noise level. These results confirm our predictions for different licensing conditions tied to individual forms we were testing – the demonstrative form appears to have strong constraints on the type of referent it can resolve to, the indefinite form is clearly very flexible with much weaker form-specific preferences, and the definite form appears to have a preference for salient, mentioned referents (unlike the indefinites), but it is a malleable constraint.

The examination of the response time data also conforms to our predictions. It reveals that responses for changes involving mentioned and non-mentioned referents show different patterns for different referential forms. Response times to accept the moved item as a referent are marginally slower for *a* N for mentioned referents, while the response times for *the* N and *that* N increase for the non-mentioned ones. This suggests that the definite and demonstrative forms prefer a mentioned salient referent and an attempt to resolve the reference to a lower salience, non-mentioned referent causes a difficulty in processing. The situation is reversed for the indefinite form. Participants are faster to identify the referent when it is of
lower salience and not previously mentioned, and are slower when the intended referent has been previously mentioned. Furthermore, participants are slower in the N and that N conditions when their response is a ‘match’ than when it is a ‘mismatch’ on trials with non-mentioned referent changes – indicating that it is easier to reject a non-mentioned referent than it is to accept it as the referent.

Examining the fixation proportion graphs in Figures 4.7 and 4.8, we replicate our basic findings from previous experiments. Increasing intervening material affects the acceptability of different referents for different anaphoric forms. We also find the most fixations to the mentioned referent in the that N condition along with a faster rise to target and a significantly extended drop-off of these fixations in the non-mentioned item moving condition (Figure 4.8, that N condition). It is possible that just as the indefinite form tends to be interpreted as ‘any’, the demonstrative condition has a special lexically specified requirement to refer back to an entity established within the discourse.

Interestingly, we do not find an effect of the discourse manipulation in this setup – the pattern of the results for different referential forms is the same with or without intervening discourse material. This is most likely a result of the different task used here compared to Experiments 1-3. Acceptability judgements task coupled with a more loose referent interpretation appear to have shifted the weight of the referent preference from discourse salience which changes with intervening discourse material, to individual form-specific licensing constraints that are independent from discourse changes.
**Non-linguistic assessment of discourse salience**

We have discussed earlier the circularity of discourse salience definition – when salience of an entity introduced by the linguistic antecedent is used to explain the felicity of different anaphoric forms, while at the same time the preferred anaphoric form is used as a diagnostic for salience. For example, a pronoun refers to the most salient entity in the discourse, but the reason it is the most salient entity is because we can refer to it with a pronoun. One approach to breaking this circularity could be via assessing discourse salience through non-linguistic means.

We have defined salience in terms of an entity’s retrievability from memory. Every mental representation consists of multiple features associated with each other – linguistic, visual, contextual, etc. We know that mentioning an object in discourse activates its visual representation (Dahan & Tanenhaus, 2005), and making an object visually salient makes it a more likely target of a produced utterance (Gleitman, January, Nappa & Trueswell, 2007). The contribution to discourse salience imparted by visual features will inevitably be different from that of linguistic properties, but they all affect each other by being related to the same discourse entity (for an overview see Myachykov, 2007; Myachykov et al., 2007). It is therefore plausible to suggest that varying discourse salience will affect the representation of the entity’s visual features and vice versa.
On some trials in Experiment 4 we introduced a change in the color of one of the displayed objects following a brief masking of the scene using a change-blindness paradigm (see Figure 4.9) (Simons & Levin, 1997; Rensink et al., 1997). We know from existing research that this type of change is normally not consciously perceived by participants. This makes detection of these changes an implicit task – as compared with the explicit referent choice from Experiments 1-3 or acceptability judgements in Experiment 4. The goal of this manipulation is to see whether such subtle changes can affect the process of reference resolution and whether a linguistically-independent assessment of discourse entities’ salience would be possible.
Figure 4.9: Sample Display – Color Change, Experiment 4. Solid lines denote movements of mentioned objects. The dotted line denotes a possible movement of a non-mentioned competitor object during the test sentence. The color change (change blindness) condition is built on top of the ‘move’ intervening condition – the object mentioned in the setup sentence changes color immediately following both the intervening sentence and the scene change corresponding to the intervening sentence. The object move is explicit, while the color change is implicit. In the initial display (first scene pictured), the bowls are red and blue, the rolling pins are red and blue, the wine glass is blue, and the cactus is green. In the color change condition, the mentioned bowl (which is now on the table) changes color from blue to green following the move by the cactus to the cupboard.
The experimental setup contained 18 color-change trials per subject – nine for each of the two types of experimental object moves (mentioned & non-mentioned). These were distributed throughout the experiment to minimize the possibility of participants noticing these changes. The change was a change in the color of the Theme of the setup sentence (the item which would then be referred to in the test sentence). All of these color change trials involved an intervening sentence, and the change always happened 200 ms after the display change following this intervening sentence – a grey mask would cover the whole screen, after which the display would return with the color change already in it. In our setup, the mask duration was set to 100ms to correspond closely both to the duration of the attention capturing manipulation in the ‘flash’ condition and to the change blindness paradigm implemented in Rensink et al. (1997). Due to its transient nature, participants usually reported this manipulation as a flicker of the screen, but without awareness of any change in the display.

In terms of predictions for the color change condition, we expect that the interference of the color change will clash with the initial mental representation of the Theme thereby drawing attention to the object and increasing its processing cost. Because of that, we expect a greater proportion of eye fixations on this item compared with trials without a color change present. There are two possibilities with respect to the referent acceptability judgements. One is that the salience of the changed object would receive a boost as a result of the clash. This will lead to an increased acceptability of the mentioned item as a referent. Alternatively, the
existing representation could be disrupted in which case it would become a less acceptable candidate for a referent resulting in a smaller percentage of match judgements.

Results

We first examine the proportions of fixations to the color change target item in the change blindness condition compared with the no color change condition (Figure 4.10). The target item is the Theme in the setup sentence, which had its color changed following the intervening material sentence. We have separated cases with different anaphors into separate graphs. The zero point represents the last scene change onset following the test sentence. Note that within 200 ms of the change we observe a sharp drop in fixations caused by the eye being captured by the moving object, which is consistent with when we would expect to see signal driven fixations. Analysis of the fixation proportions demonstrates a greater proportion of fixations to the changed item in the color change condition for all referential forms. These differences reach significance for the indefinite and the demonstrative conditions (see area under the curve analysis below).
Figure 4.10: Fixation Proportions during Color Change Trials. Proportion of fixations over time to the color change target comparing color change trials with trials without a color change, grouped by the referential form used in the test sentence (that N, the N, a(n) N). The graphs are aligned at the display change onset (following the test sentence).
**Figure 4.11: Mean Fixation Proportions.** Mean area under the curve of fixation proportions, aligned at the display change onset (Figure 4.10), for the time window –1600 ms to 0ms.
Figure 4.12: Percent of ‘Match’ Responses on Color Change Trials. Proportion of responses indicating a match between the scene change and Lisa’s actions in the color change trials.

Figure 4.13: RTs of ‘Match’ Responses on Color Change Trials. Response times for color change trials in which participants gave a ‘match’ response.
To analyze the magnitude of this effect, we calculated a mean area under the curve for the time window –1600 ms to 0ms. 1600 ms approximates the average distance between the determiner (the point when participants can start to identify the referent) and the change onset, which is 1636 ms. The results are displayed in Figure 4.11, which presents mean fixation proportions for all referential forms. A repeated measures ANOVA on data transformed as in the previous experiments revealed a significant main effect of the referential form, $F(1,15)= 33.188$, $p < 0.001$, and of the trial type (color change or no change), $F(1,15)= 20.644$, $p < 0.001$, as well as an interaction, $F(1,15)= 6.354$, $p = 0.024$. The magnitude of the fixation proportions is the smallest in the $a$ N condition, greater in the $the$ N, and the greatest in that N. The proportion of fixations is significantly greater for the color change condition compared to the trials without the color change for the indefinite form, $F(1,15)= 10.336$, $p = 0.006$, and the demonstrative form, $F(1,15)= 29.591$, $p < 0.001$. The greatest increase in fixations as a result of the color change occurred in the ‘that N’ condition – it is significantly greater than the ‘a N’ ($F(1,15)= 6.306$, $p = 0.024$) and the ‘the N’ ($F(1,15)= 9.334$, $p = 0.008$) conditions.

We now look at the acceptability judgements data. Figure 4.12 presents the percentage of ‘match’ responses for the three anaphoric forms across different types of display changes when: (a) a mentioned item moved; (b) a non-mentioned item moved; and (c) a mentioned item moved to a wrong location, which serves as a baseline error-rate comparison. A repeated measures ANOVA performed on the transformed data revealed significant main effects of anaphoric form, $F(1,15)= $
62.741, p < 0.001, and of the type of scene change, F(1,15)= 105.190, p < 0.001, and an interaction, F(1,15)= 93.994, p < 0.001. Since all color change trials belong to the ‘move’ condition, we compare the mentioned and the non-mentioned conditions to their respective counterparts in Figures 4.2 and 4.3. The comparison demonstrates no significant differences in the proportion of ‘match’ responses between the color change trials and trials without any color changes present.

Comparing the response times in the color change condition (Figure 4.13) to the response times in the no change trials (Figures 4.4 and 4.5), we observe the same results as before – the indefinite form takes longer in the mentioned move condition than the definite and the demonstrative forms, with this pattern reversed for the non-mentioned move condition. The overall levels of the response times are also similar to prior results.

**Discussion**

Our investigation of the proportions of fixations on the changed item revealed a significantly greater proportion of fixations in the color change condition (90 vs 60 average mean area under the curve, Figure 4.11). This suggests that the object color was at least partially encoded in the mental representation formed by the discourse, and our participants were implicitly detecting the change without being aware of it on a conscious level. The greatest increase in fixations as a result of the color change occurred in the ‘that N’ condition. One possible explanation is that the demonstrative
form is interpreted contrastively and thus draws attention to any differences between the mentioned and the non-mentioned entities.

Analysis of the color change acceptability judgements and response times reveals no significant differences between the color change and the no-color-change conditions. Overall, we can say that the implicit sensitivity to the manipulation (demonstrated by the fixation proportions data) failed to affect the explicit judgments. This result is nevertheless not inconsistent with our hypothesis that the manipulation would increase an entity’s salience thereby making it more acceptable as a referent, not less as a disruption of the representation would. The reason we do not find an increased rate of ‘match’ judgements for the color change condition is probably the fact that our data is too close to ceiling.

An important result to note is the differential effect of the color change manipulation on different referential forms – greatest on *that N*, smallest on *the N*. Such differential effects can be used as a possible diagnostic tool for discourse salience. Overall, although the effects of the color change manipulation are suggestive, they are sufficiently promising to suggest that the paradigm might be useful as a tool for establishing an independent, non-linguistically motivated measure of discourse salience.
Chapter 5 – Summary and Conclusions

The goal of the current experiments was to investigate the validity of predictions from the Givenness Hierarchy by Gundel et al. (1993) and similar theories that associate different levels of discourse entities’ salience or different cognitive states with various anaphoric forms (Ariel, 2001). This salience-based approach has been challenged over the years with alternative frameworks such as work by Kaiser (2003), Aissen (1999) and Hofmeister (2007). In particular, we examine the degree to which form-specific constraints on referential form use and interpretation can override salience-based preferences. We also address a long-standing question of defining discourse salience in functional terms.

We utilized the visual world paradigm (Tanenhaus et al., 1995) to assess the reference resolution process in an online setting. The use of object displays and schematic scenes allowed us to manipulate the initial salience level of discourse entities and, through discourse manipulations, to examine how discourse salience changes, and whether it is affected by factors other than the way the antecedent was introduced.

In Experiments 1 and 2 we introduced a methodology in which we used changes in the unfolding discourse to change salience of the discourse entities. This approach allowed us to assess the effect of these changes in salience on the three different anaphoric forms we selected for these studies (that N, the N, a(n) N). These referential forms belong to different levels of the Givenness Hierarchy with
progressively lower hypothesized salience of their referents. The referential forms were affected differently, with increasing intervening material having the strongest effects on indefinites and the weakest effect on the demonstratives. The initial preference across all forms was for the discourse mentioned entity. The preference of the indefinite form for mentioned referents is inconsistent with the assumptions of the Givenness Hierarchy that the indefinite form cannot be used to felicitously refer to a highly salient discourse entity.

Experiment 3 investigated possible reasons for the preference of indefinites for the mentioned referents. One alternative was that the indefinite form is interpreted as consistent with an ‘any’ interpretation. In this case, the salience of the alternative referents is not the main deciding factor, and the choice of the referential expression would be largely specified by the form-specific licensing conditions. The other possibility is that the indefinite form requires a choice from among several entities of roughly the same salience level and when this is not possible, the felicity conditions of its use are being violated. The experiment discriminates between these alternatives by adding a third exemplar of the type of entity referred to in the discourse setup sentence. This created a situation where there is both a salient mentioned entity and two non-mentioned alternatives of equal salience which are less salient than the mentioned entity. The results of this experiment largely replicated the results from Experiments 1 and 2, suggesting that discourse salience levels do not significantly affect the use and interpretation of the indefinite anaphor.
In Experiment 4 we investigated the same questions with more structured scenes, a narrative discourse, and a scene verification task. Importantly, we used dynamic scene changes during the narrative description to manipulate the alternatives participants had to consider in making their response. Participants made judgements of the goodness of fit of a referent to a referential form based on whether the scene change matched its description. Again, the results largely matched our previous findings. We found an overall preference for salient mentioned referents, which was strongest for ‘that N’ and somewhat weaker for ‘the N’. The indefinite form resolved equally well to mentioned and non-mentioned referents consistent with an ‘any’ interpretation. Participants did not accept a non-mentioned entity as a possible referent for a demonstrative reference, while the acceptance rate for the definite reference was about 40%.

Experiment 4 also contained scene manipulations intended to evaluate the relative salience of discourse entities directly. We introduced a small proportion of trials with a color change manipulation using a change blindness paradigm (Rensink et al., 1997), which probed the mental representation of a potential referent by changing its color. The assumption was that this change would clash with the existing mental representation and interfere with the entity’s retrievability from memory. Although fixation proportions indicated that the manipulation was successful, with participants registering the change on an implicit level, the manipulation failed to affect the explicit judgments pattern. However, from the implicit change detection data we found that the change was detected differently
under different anaphoric conditions which suggests that it can be used as a
diagnostic. This manipulation demonstrated the feasibility of monitoring discourse
representations through non-linguistic means. It will be particularly interesting to
evaluate whether introducing new discourse entities or shifting the focus of the
discourse will affect participants’ ability to detect color changes for a mentioned
entity. It will also be important to determine whether color changes are better
detected for entities introduced in higher-ranked thematic roles compared to entities
introduced in lower-ranking thematic roles.

The results reported here have clear theoretical implications for models of
reference resolution and theories of discourse salience. Our findings are problematic
for approaches that use a single dimension, such as salience or accessibility to
associate different referring expressions with their preferred potential referents. The
framework that best accommodates our results is the form-specific multiple-
constraints approach advocated by Kaiser and colleagues (Kaiser & Trueswell, 2008)
which proposes that while an entity’s salience is important for determining the
appropriate reference, individual anaphoric forms carry their own specific constraints
on the conditions of their felicitous use.

Our results also have methodological implications. We introduced a change
blindness manipulation to assess the representation of an entity in memory. Although
the results were suggestive, they demonstrated the feasibility of assessing discourse
representations without relying on linguistic manipulations. In the future it will be
important to explore how the paradigm can be extended to different types of changes
in order to further explore the nature of the representation encoding for a discourse entity and how that representation is affected by different referring expressions and circumstances of their use.
References


Jaeger, T. F. (in press). Better categorical data analysis: Away from ANOVAs (transformation or not) and towards Logit Mixed Models. Journal of Memory and Language. (Special issue on Emerging Data Analysis).


Rensink, R.A., O’Regan, J.K., & Clark, J.J. (1997). To see or not to see: The need for attention to perceive changes in scenes. Psychological Science, 8, 368–373.


