Tracking eye movements to uncover the nature of visual-linguistic interaction in static and dynamic scenes

Caroline van de Velde

A Thesis

in

The Department

of

Psychology

Presented in Partial Fulfillment of the Requirements
for the Degree of Doctorate of Philosophy (Clinical Psychology)
Concordia University
Montréal, Québec, Canada

April 2008

Your file Votre référence
ISBN: 978-0-494-37735-2
Our file Notre référence
ISBN: 978-0-494-37735-2

ABSTRACT

Tracking eye movements to uncover the nature of visual-linguistic interaction in static and dynamic scenes

Caroline van de Velde, Ph.D.
Concordia University, 2008

These studies examined the role of sentence and visual context in the access to verb-complement information, using a new eye tracker and change blindness paradigm. Participants' eye movements were monitored as they viewed pictures of objects (Experiment 1) or dynamic scenes (Experiment 2), and listened to related sentences. In Experiment 1, two sets of sentences were contrasted, a highly constraining causative construction in which there was a close conceptual relation between the verb and its direct object (e.g., "The woman *burned* the *candle*") and a neutral construction with a transitive perception verb (e.g., "The woman *admired* the *candle*"). Starting at three different points within the presentation of the verb (onset, middle, offset) and noun (onset, offset, offset+200 ms), participants' task was to look and indicate whether the target objects mentioned in the sentences (e.g., "*candle*") were present in the visual displays. Results indicate that semantic information extracted at the verb can be used to constrain the domain of reference in the scene and in some cases predict the referent of the grammatical complement of the verb, depending on tasks demands, conceptual consolidation of the scene, and the presence of competitor objects.

In Experiment 2, two different classes of verbs were contrasted, denominal and non-denominal verbs, which either implicitly (e.g., "The woman will <u>*iron*</u> the *shirt*") or explicitly (e.g., "The woman will *chop* the *vegetables* with the <u>*knife*</u>") named the instrument nouns. These movies were edited, unbeknownst to the participants, so that the

real referents of the verb's grammatical object (e.g., "*shirt/vegetables*") or instrument (e.g., "*iron/knife*") gradually dissolved. Participants' task was to provide a detailed description of each scene, and later perform a recognition task. Although results indicate that information extracted from the sentence helped identify, describe and remember scene details, visual context seems to take precedence over linguistic input properties in guiding eye movements. In conclusion, the processor appears to incrementally integrate all available knowledge, linguistic and non-linguistic, with the aim to rapidly interpret the linguistic description of what we see in the world and how we may interact with it.

Acknowledgments

First and foremost, I am indebted to my supervisor, Dr. Michael von Grünau, for sparking my interest in research, as well as for his continuous support, guidance and friendship throughout the years. My gratitude also extends to Dr. Roberto de Almeida for his priceless insight and assistance in this process, as well as his contagious passion and enthusiasm for research. In addition, I am thankful to Dr. Rick Gurnsey for accepting to venture into a new area of research so that I could benefit from his valuable thoroughness and critical thinking. Last but not least, I am grateful to Dr. Pierre Jolicoeur and Dr. Paul Albert for their insightful comments and suggestions on an earlier manuscript.

I am thankful to the friendly staff at Ikea for letting us film at their store and allowing us to freely use their decorative rooms and merchandise to produce realistic movie scenes. I am grateful to Robert Wilson for his professional efficiency in producing the short movies used for this study and Bisser Maximov for his great talent, patience and diligence in editing them. I would also like to thank Nancy Wada, Alexandre Pepin, Angela Vavassis, Jeff Yuen, and Alla Sorokina for their award-winning acting performances. For their help in recruiting and running participants, I would like to thank Zinnia Madon, Jake Godfrey, Mike Belance, Heather Wilcox, and Sally Cooper. Finally to my fellow colleagues and the undergraduate students who participated in all these studies, thank you for lending me your eyes and ears to gain some insight into the workings of the brain.

I am forever grateful to my parents, family and close friends, without whose continuous love, support, and encouragement I would not have been able to attain this important goal. I am especially indebted to my partner, for being my personal cheerleader throughout this long journey and providing me with the emotional and financial peace of mind to realize my dreams.

Last but not least, I am grateful for the financial support I received, in the form of graduate fellowships from FQRC, Concordia University, and generous donations from the Gold and Renaud families. I am also thankful for the many opportunities to share my work with the scientific community at large through the financial support of departmental travel grants, as well as Dr. von Grünau's and Dr. de Almeida's research grants.

Table of Contents

List of Figures

List of Tables

Tracking eye movements to uncover the nature of visual-linguistic interaction in static and dynamic scenes

Introduction

The eye has often been compared to a camera. Light comes through the cornea, a transparent cover that is like the glass of the camera's aperture (see Figure 1). The amount of light that enters the eye is controlled by the pupil, an opening that dilates and contracts like the shutter of a camera. Similar to the lens of a camera, the light that enters the pupil is focused onto the retina (a lining of light sensitive cells on the back of the eye) by a biconvex crystalline lens. The retina then acts like the film in a camera by reacting to the incoming light and sending the signal to the brain via the optic nerve. After this point, however, the two systems diverge with regards to how the signal is developed into its final product. Some examples of how a photograph differs from a mental image include: (1) a non-edited sharp photograph is usually an accurate/reliable representation of its subject, without blurriness, additions, subtractions or changes made from the time it was taken to when it was produced; (2) taking simultaneous photographs of the same subject using the same brand of camera will produce identical pictures; (3) with passing time, the content of well-preserved photographs will not change; and (4) after clicking the camera button, the photograph will not be affected by extraneous factors. In human visual perception, however, the act of seeing is a constructive and interactive process constantly influenced by low and high-level processes (e.g., color, attention, world knowledge, expectations, linguistic input, etc.); limited by the capacities of short and long-term memory; as well as the system's fallibility in certain circumstances (e.g., illusions, failure to detect changes, etc.).

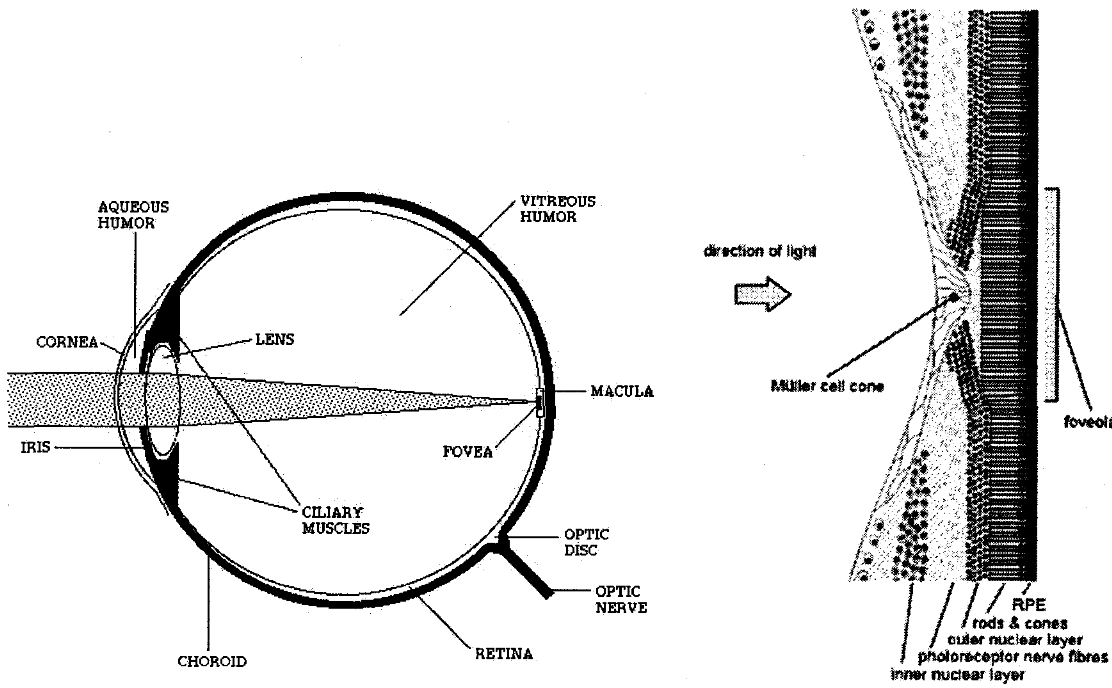


Figure 1. Schematic representation of the eye and fovea.

These fundamental issues will be discussed in further detail, as they provide an important backdrop to the set of experiments that will be addressed in this paper. More specifically, eye movements were recorded to investigate how the brain interprets visual and linguistic information in real-time. By monitoring the timing and pattern of eye movements across static and dynamic scenes, a better understanding of how these two cognitive systems interact with each other during sentence comprehension was achieved. These studies, in particular, explored different verb classes (e.g., causative, perception and denominal verbs) and how they directed visual attention (as measured by eye movement responses) to the referents of their grammatical objects and instruments in scenes. The results are then discussed in the context of a model which proposes how the linguistic and visual systems interact to help us make sense of what we see and hear in context.

Eye Movements and Visual Perception

The eye's retina is composed of two types of sensors called the rods and cones (see Figure 1). These two sensors differ in their visual sensitivity and their distribution across the retina. Rods which are light-sensitive cells are responsible for vision in low illumination conditions. When exposed to high levels of luminance the rods become saturated and only the cone photoreceptors remain functional. In high illumination conditions these cones are responsible for the perception of fine detail and color. In the periphery, the rods greatly outnumber the cones, which allows for night vision but leads to poorer visual acuity. On the other hand at the center of the eye of the macula is a small pit containing a high concentration of cone photoreceptors, referred to as the fovea. Compared to the rest of the retina, the cones in the fovea are smaller in diameter and are, therefore, more densely packed. The high spatial density of these cones accounts for the high visual acuity capability at the fovea. The fovea sees only the central two degrees of the visual field, which is roughly equivalent to twice the width of your thumbnail at arm's length (Fairchild, 1998).

In order to efficiently use the greater resolution of the fovea, the human visual system continuously reorients the fixation point to different regions of the visual environment. These reorientations occur on average three times each second via saccadic eye movements (Henderson & Hollingworth, 1999a). A saccade, which is the fastest movement of an external part of the human body, can reach up to velocities of 900 degrees per second (Carpenter, 1988). A saccade during scene perception can last between 50 to 1000 milliseconds (ms), with an average of about 330 ms (Henderson & Hollingworth, 1998). In order to avoid blurring of the image on the retina during a saccade the visual system has a mechanism which essentially suppresses retinal processing (Volkman, 1976). Between any two saccades is a period of fixation during which the eyes are relatively stationary, and when virtually all visual input occurs (Matin, 1974). The active combination of head and eye positioning provides humans with the constant illusion of high resolution vision.

Eye Movements and Scene Perception

Many studies throughout the years have examined the processes involved in controlling fixations during scene perception. According to Henderson and Hollingworth (1999a) a scene is typically defined as a:

> "...semantically coherent (and often nameable) view of a real-world environment comprising background elements and multiple discrete objects arranged in a

> spatially licensed manner. Background elements are taken to be large-scale, immovable surfaces and structures, whereas objects are smaller-scale discrete entities that are manipulable within the scene."

For instance an encompassing view of a kitchen would be considered a scene, with the floor and walls as background and the vegetables and knife as objects within the scene. In particular, studies have attempted to understand where fixations will be made in a scene, and subsequently for how long they will remain paused in these particular locations. In 1935, Guy T. Buswell was the first to methodically record and analyze the eye-movements of observers as they viewed complex scenes. In this investigation the eye movements of 200 participants were recorded while they viewed 55 photographs of different types of artwork (e.g., paintings, sculptures, tapestries and architecture). When examining viewer's fixation patterns for each image Buswell (1935) observed that regions rich in information were often the areas with the highest density of fixations. Moreover, when examining the eye movement data across participants he noticed that although viewers did not necessarily make fixations in the same temporal order, they tended to fixate on the same spatial locations in a picture. Buswell (1935), thus, concluded that these consistencies across participants and images suggest that humans do not randomly explore pictures. Rather, the eyes tend to focus on foreground elements, such as people and faces (features rich in non-verbal communication and information), rather than background elements, such as clouds or trees (features poor in information).

Yarbus (1967), another great pioneer in the study of eye movements in complex images, asked viewers to examine IE Repin's color painting "An unexpected Visitor" and follow a set of seven instructions (see Figure 2). Yarbus (1967) found that viewer's

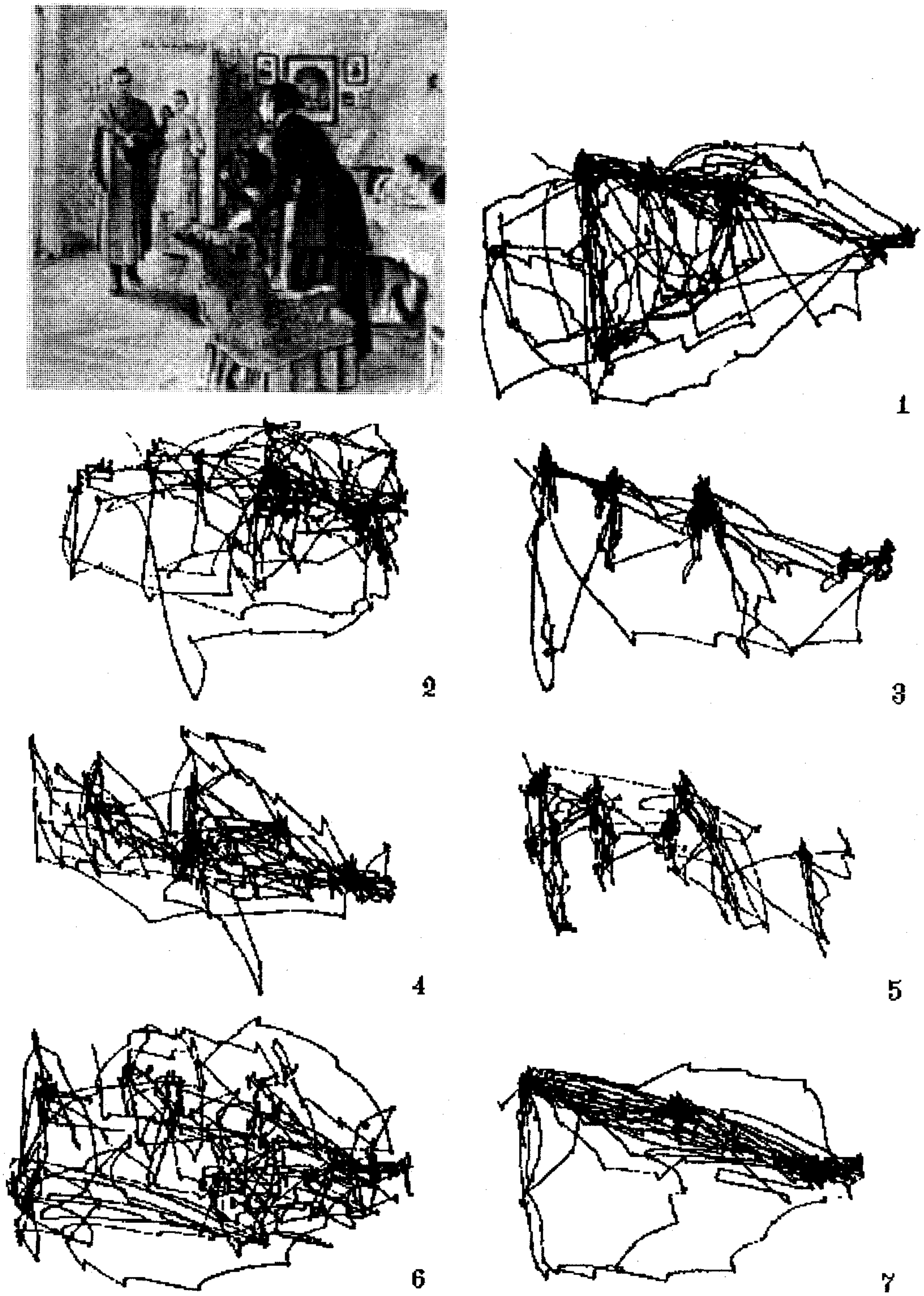


Figure 2. Eye movement data by one subject for each task (Yarbus, 1967): 1) Free examination; 2) estimate the wealth of the family; 3) give the ages of the people; 4) surmise what the family had been doing before the arrival of the "unexpected visitor"; 5) memorize the clothes worn by the people in the room; 6) memorize the location of the people and objects in the painting; and 7) estimate how long the "unexpected visitor" had been away.

pattern of eye movements and fixations varied dramatically depending on the type of instructions given. For instance in task 3, in which they were asked to determine the ages of the people in the scene, viewers tended to concentrate most of their fixations on the people, in particular their faces. On the other hand, in task 6, in which viewers were asked to recall the location of the people and objects in the painting, fixations tended to be broadly distributed over the scene. Based on these results and the eye movement patterns obtained from other types of pictures (e.g., portraits), Yarbus (1967) concluded that the eyes were not merely reflexively attracted to the physical features of an image, but rather tended to be directed towards the areas in a picture which contained the most "useful or essential" information for the task at hand.

Several studies since Buswell (1935) and Yarbus (1967) have replicated the finding that more informative scene regions have higher fixation density (Mackworth & Morandi, 1967; Pollack & Spence, 1968; Mackworth & Bruner, 1970; Antes, 1974). Yet, a question remains: What is it about an area that causes it to be rated more informative and, therefore, fixated more often than other areas? Are these ratings based on visual (bottom-up) or cognitive (top-down) factors or a combination of both? Loftus and Mackworth (1978) were the first to examine whether semantic informativeness influenced fixations when visual informativeness was held constant. They defined an object as being semantically informative if it had a low probability of being in the picture, based on the gist of the scene and the observer's past history with such a scene. For instance, as shown in Figure 3, the tractor in the top panel was considered to be a non-informative object because it had a high probability of appearing in a farm scene. By contrast, the octopus in the bottom panel of Figure 3 had a low probability of being

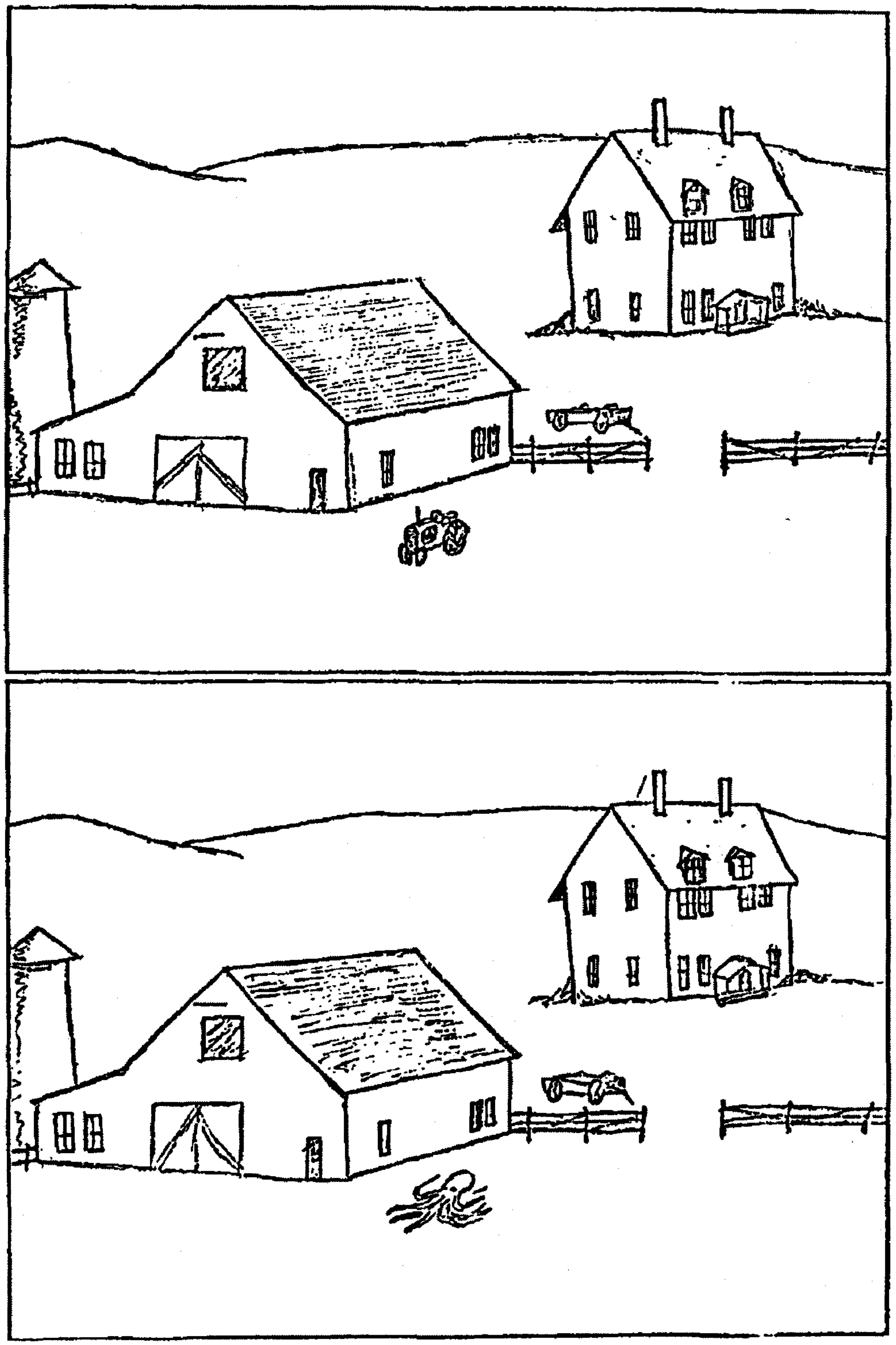


Figure 3. A picture used in a study by Loftus & Mackworth (1978). In the top panel is an example of a scene with a "noninformative" object (the tractor), and the bottom panel the same scene but with an "informative" object (the octopus).

found in a farm scene and was, therefore, considered to be an informative object. Participants' eye movements were recorded as they viewed 78 pictures and instructed that they would later be asked to recognize each picture. The results of this study indicated that participants tended to fixate the informative objects (e.g., octopus) earlier, more often, and for longer durations than the non-informative objects (e.g., tractor). Loftus and Mackworth (1978), therefore, concluded that the semantic information visually acquired from the periphery (the average distance to the target object was over 7° of visual angle) controlled the initial placement of a fixation in a picture.

Although several studies have since replicated the finding that semantic informative regions tend to attract more fixations and re-fixations, they have failed to replicate Loftus and Mackworth's (1978) finding that fixations are drawn earlier to these regions (De Graef, Christiaens & d'Ydewalle, 1990; Henderson, Weeks, & Hollingworth, 1999; Mannan, Ruddock, & Wooding, 1995). In fact, in a study by Henderson et al., (1999), the opposite pattern of results was obtained using a visual search task. Henderson and colleagues asked participants to view 24 complex line drawings and search for a target object in each scene. The target object was cued by a visually presented word before each trial. Results demonstrated that viewers found the scene consistent target objects more quickly and fixated them earlier, than the inconsistent scene items. Henderson et al. (1999) argued that participants were able to use their knowledge about the likely position of consistent objects to drive their eye movements to the target objects (e.g., a toaster is usually found on a flat surface such as a counter). By contrast, in experiment 1 in which participants were instructed that they would later receive a memory test, no such differences were found. Henderson et al. (1999) hypothesized that this was the result of task-specific strategies, such that in a memory test participants may not have been as motivated to quickly fixate the inconsistent objects, as compared to when asked to perform a time-pressured visual search task.

Henderson et al. (1999) offered several explanations to account for the incongruent results between studies. For instance, they suggested that the shorter viewing time (4 sec.) in Loftus and Mackworth's study (1978) may have motivated participants to fixate the inconsistent objects more quickly. Another explanation involved the type of stimuli used for each study. More specifically, they argued that the scenes used in the follow-up studies may have been more complex than the ones employed by Loftus and Mackworth (1978), making peripheral semantic analysis more difficult. Line drawings provide the added benefit of eliminating the effects of color, luminance, and contrast. On the other hand, as shown in Figure 4, by attempting to make line drawings more complex and realistic (line drawings derived directly from photographs), all the lines from the foreground and background objects make the target objects less discernable, especially at exposure times of 8 and 15 seconds. In fact in the study by Henderson et al. (1999), participants reported finding it "difficult to evaluate the semantic consistency between an object and its scene until they had fixated relatively close to that object." Yet, according to the results of another study by Henderson and Hollingworth (1998), in which fixation durations were compared between full-color photographs, full-color renderings of scenes, and black-and-white drawings, only small differences in mean fixation durations were found across conditions – with slightly longer fixation durations for the full-color photographs.

Visual factors, on the other hand, appear to play an important role in extracting

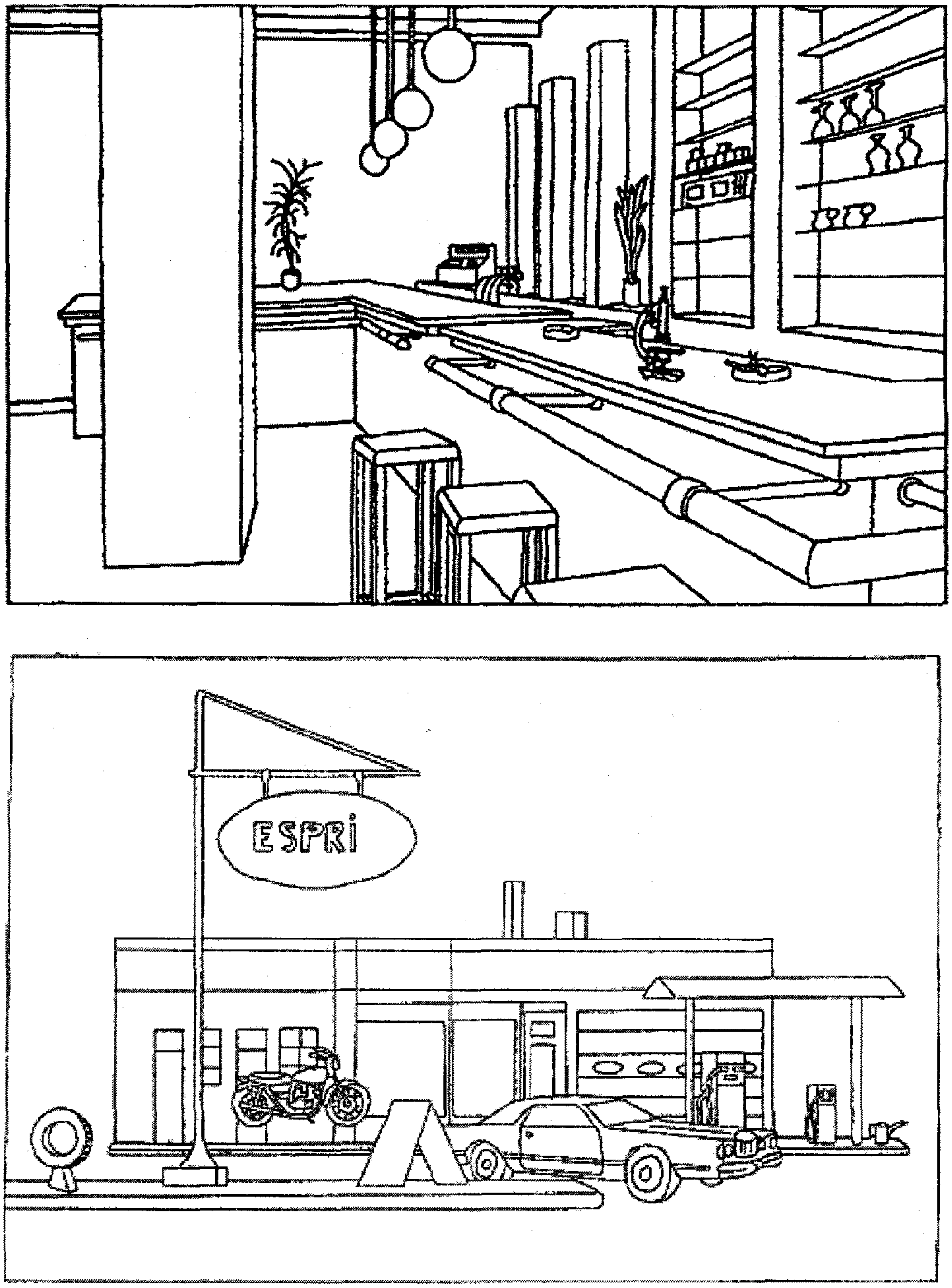


Figure 4. Top panel is a sample stimuli used by Henderson et al. (1999a), and the inconsistent object is the microscope on the bar. Bottom panel is a sample stimuli used by De Graef et al. (1990) and the inconsistent objects are the motorcycle (positioning) and the gas dispenser (size).

the "gist" of a scene. Several studies have shown that scene "gist" is extracted very rapidly, occurring simultaneously or even before object identification (e.g., Potter, 1975, 1976; Biederman, 1981, 1988; Schyns & Oliva, 1994; Oliva & Schyns, 2000). Using diverse stimuli, methodologies, presentation latencies, and response measures, these studies have obtained scene identification latencies ranging between 125 and 250ms. For instance, in the series of studies by Potter (1975, 1976), slides of real-world scenes (e.g., a city, a living room, a highway, etc.) were projected in rapid succession. Participants were asked to press a button as soon as they recognized a target scene (e.g., a highway) that was verbally provided prior to the presentation of the slides. Potter found that participants were able to perform the detection task with high efficiency. Potter (1993) suggested that pictured scenes can be understood within about 100 ms, but that additional processing is required to consolidate the identified pictured scenes into short-term memory. In fact Intraub (1999) proposed that in order to minimize memory overload, transsaccadic memory (the memory for information maintained across saccades), is not highly detailed; containing only the scene's meaning and general layout, and some detail. Intraub (1999) surmised that if the visual system requires additional detail, it can simply fixate the region of interest in as little as 200 ms (Matin, Shao, & Boff, 1993).

In summary, according to Henderson and Ferreira (2004) the eyes tend to fall on objects or clusters of objects, and rarely on empty spaces between them – unless that location previously contained an object of interest (discussed further in the second series of studies). Furthermore, it seems that both visual (bottom-up) and cognitive (top-down) features interact at different stages of scene processing to control the location and duration of fixations. Within about 100 ms of viewing a scene participants have used the visual characteristics of the scene to obtain the "gist" of the scene. From the scene "gist" viewers then use their knowledge of the scene schema (e.g., computer is usually found on a flat surface such as a desk), as well as the visual features of the scene, to control their first few fixations. It is only after the viewer has become familiar with the scene, by fixating and semantically analyzing the different regions of the scene, that re-fixations become controlled by the semantic informativeness of a region. Furthermore, the needs of the viewer (i.e., task at hand), as well as the amount of pre-exposure to a scene, determine how strongly visual and semantic characteristics of a scene influence fixations. For instance, observers who are quite familiar with a particular scene (e.g., their kitchen) will rely mostly on memory of the scene to control the positioning of their fixations. In conclusion, these studies demonstrate that eye movement recordings provide us with a valuable real-time, non-invasive and sensitive measure to study high-level cognitive processing.

Eye Movements and the Field of Psycholinguistics

The proven efficacy of measuring eye movements to capture real-time cognitive processing resulted in a rapid methodological "paradigm shift" in the field of psycholinguistics. In the last 20 years this new research tool has been principally used for the detailed study of word recognition, syntactic parsing and sentence interpretation. One main issue that these studies have attempted to resolve is the nature of the architecture of the language processing system. In particular whether or not its computations are autonomous and insensitive to contextual and background information (Fodor, 1983, 2000; Garfield, 1987).

Fodor (1983), a well known advocate of the "autonomous" viewpoint, elaborated on the idea that language is "modular." According to this view, lateral flow of information is allowed between each level within the language processing system (i.e., lexical, syntactic, and semantic), but information from other cognitive processes, such as world knowledge, can not penetrate the module. An example he used to demonstrate this argument is the persistence of illusions (e.g., Mueller-Lyer Illusion – see Figure 5). Even after being aware that it is not what it appears to be the illusion does not simply disappear because we really want it to or because we know that the lines are of the same length (i.e., no top-down interaction). Fodor (1983) hypothesized that each module is domain specific and thus the computational systems that deal with the production of language, for instance, have nothing in common with those that deal with the analysis of visual perception. Moreover, Fodor (1983) proposed that processing within a module is unconscious, mandatory (e.g., can not prevent understanding a word in one's own language), fast, and low-level. Fodor (1983) argues that it is only after the modules have released their output into the central systems that the central processors integrate the different types of information (i.e., interact) from each module, along with information stored in long-term memory, to form a stable and coherent whole. Fodor (1983) hypothesized that because the central processors are slow and unable to integrate or hold in conscious awareness outputs from all modules, one of its key functions is to selectively attend to the information that is most relevant to the task at hand.

Ever since Fodor (1983) proposed the modularity model the research approach in psycholinguistics has been to investigate the degree to which language processing can proceed without the use of contextual information. The important distinction to be made between "modularists" and "interactionists" is not whether context has an effect on

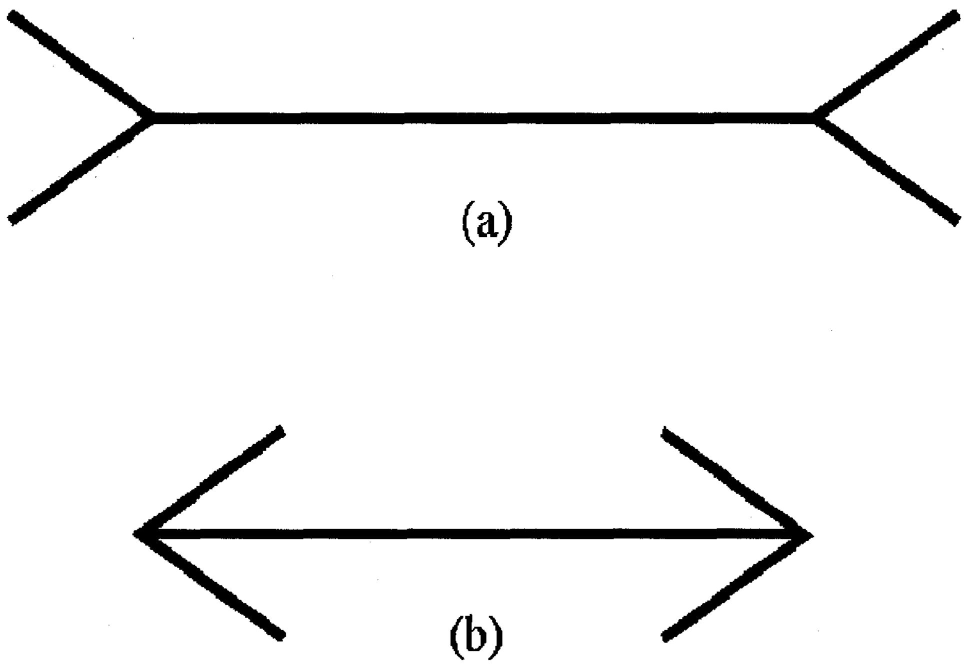


Figure 5. Schematic representation of Mueller-Lyer's optical illusion. Both set of arrows are the same size, however, when asked to judge the lengths of the two lines viewers will typically claim that (a) is longer than (b).

language processing, but rather *when* context has an effect. For instance, in sentence interpretation "interactionists" (e.g., Marslen-Wilson & Tyler, 1989) argue that other sources of information (e.g., semantic, discourse, world-knowledge, etc.) can influence parsing (i.e., the analysis of the grammatical structure of a sentence) at an early stage. On the other hand, "modularists" argue that the initial stages of parsing can only be influenced by syntactic information, and that other sources of information are instead used subsequently to evaluate and if necessary revise the initial syntactic representation. Tanenhaus, Carlson and Trueswell (1989) argue that although models regarding the nature of the mind's architecture are often dichotomously categorized, a lack of consensus amongst members within the same group suggests that models fall along a continuum. According to this view, completely "encapsulated" and "interactive" models are situated at the extreme ends of the continuum, whereas the restricted interactive models fall somewhere in between (Tanenhaus et al., 1989).

In previous years, the use of discourse context (e.g., Altmann & Steedman, 1988; Ferreira & Clifton, 1986) in printed text had been used as the critical test for the interactive or autonomous character of the language processing system. In these studies, scenarios were set up using printed text (i.e., discourse context) and reading time was measured at a point within the sentence that contained local ambiguity. For instance, in the following sentence used by Altmann and Steedman (1988), "The burglar opened the door with the faulty lock and quickly slipped inside", the critical element is the word "faulty." Because according to Frazier's (1989) minimal attachment theory, which states that sentences are interpreted in the structurally simplest manner, readers will attach "with" to the verb phrase (i.e., "opened") as the instrument to carry out the action. Thus, when the unexpected word "faulty" is reached the system realizes that the analysis is wrong and returns to reanalyze the sentence (i.e., "with" will then attach to the noun phrase "the door"). In the study by Altmann and Steedman (1988), however, participants received one of the following contexts before the presentation of the target sentence: [1] "A burglar was trying to break into a house and had a credit card in his hand. He wanted to open a door. He found a door which had a faulty lock and a window which had a cracked frame"; or [2] "...He found a door which had a faulty lock and a door which had a cracked frame." Altmann and Steedman (1988) found that participants were faster at reading the ambiguous target sentence within context [2], because the mention of two doors resulted in participants correctly interpreting "with" as the modifier of which door will be opened – thus preventing readers from going down "the garden-path." Altmann and Steedman (1988), therefore, concluded that as sentences unfold, word-by-word, the language processor constructs alternative syntactic interpretations in parallel and then uses discourse context to immediately disambiguate them. This is often labeled as a "weak modularity" model (Altmann & Steedman, 1988).

Although the results of reading studies suggest that linguistic representations and contextual information interact at some point during language comprehension, questions remain as to *when* and *how* they do. These questions remain unanswered in part because of the inherent limitations of reading studies (e.g., restrictive reading measures, context being ill-defined and possibly weaker in reading as they require memory retrieval). Consequently, within the field of language processing research a division occurred with two main approaches emerging – labeled by Clark (1992) as the "language-as-action" and "language-as-product" approaches. In the "language-as action" approach, spoken language processing is studied using off-line methods (i.e., response measures are not time-locked to the linguistic input), such as participants engaging in interactive and cooperative dialogue during natural tasks, generally in settings with real-world referents and well-defined behavioral goals. For example, in Clark and Wilkes-Gibbs's (1986) study, a "matcher" rearranged shapes on a grid to match the hidden arrangement of the "director's" grid via collaborative dialogue. In these studies context includes the time and location of the conversation, the collaborative process between the speaker and listener, as well as their conversational goals. By contrast, in the "language-as-product" approach, language processing is studied in real-time, using fine-grained reaction time measures that are closely time-locked with the unfolding linguistic input. For instance, in Meyer and Schvaneveldt's (1971) study, reaction times were measured for target words when they were preceded by semantically related or unrelated prime words. In these studies lexical, syntactic, semantic, discourse and conversational context are carefully manipulated in order to examine how and when they affect linguistic processing (see review by Spivey, Tanenhaus, Eberhard, & Sedivy, 2002; Trueswell, Tanenhaus, & Garnsey, 1994).

The Use of Visual Context in Language Comprehension

Tanenhaus and colleagues have attempted to bridge the gap between the "language-as-action" and "language-as-product" approaches, by taking key elements of each and combining them into one innovative technique – now formally referred to as the "visual world paradigm" (e.g., Tanenhaus, Spivey-Knowlton, Eberhard & Sedivy, 1995; Eberhard, Spivey-Knowlton, Sedivy & Tanenhaus, 1995; Sedivy, Tanenhaus, Chambers & Carlson, 1999; Altmann & Kamide, 1999; Spivey, Tanenhaus, Eberhard & Sedivy, 2002; Chambers, Tanenhaus, Eberhard, Filip, & Carlson, 2002; Kamide, Altmann, & Haywood, 2003; and Huettig & Altmann, 2005). This new technique involves monitoring participant's eye-movements while following spoken instructions to manipulate objects in the visual world. Using this technique, they have been able to examine referent context effects on syntactic ambiguity while performing tasks with clear behavioral goals (i.e., language-as-action approaches), using a temporally precise measure that allowed for a moment-by-moment look at language comprehension (i.e., language-as-product approaches). The inspiration for this methodology was the result of a pioneering study by Cooper (1974) which demonstrated that listener's eye movements towards objects in a visual array tended to be closely time-locked with objects referred to in a pre-recorded text. The refinements later made to this technique by Tannenhaus and colleagues represented a significant advance in the way we study the basic cognitive and linguistic factors involved in the process of language comprehension.

One such key study, (Tanenhaus, et al., 1995 – described in greater details by Eberhard et al., 1995) involved administering participants spoken commands to interact with objects (e.g., blocks, playing cards) on a table. In the first experiment the identity of the blocks was manipulated (by changing the markings, color and shape) to vary the point at which the referential expression became unambiguous. For example, the display shown in Figure 6 would be accompanied by the instruction "*Touch the starred yellow square*." They hypothesized that disambiguation will occur at the first adjective (*starred*) for the early condition, the second adjective (*yellow*) in the mid condition, and the final noun (*square*) in the late condition. An analysis of the eye movements indicated that participants processed the instructions incrementally, rather than waiting until the end of

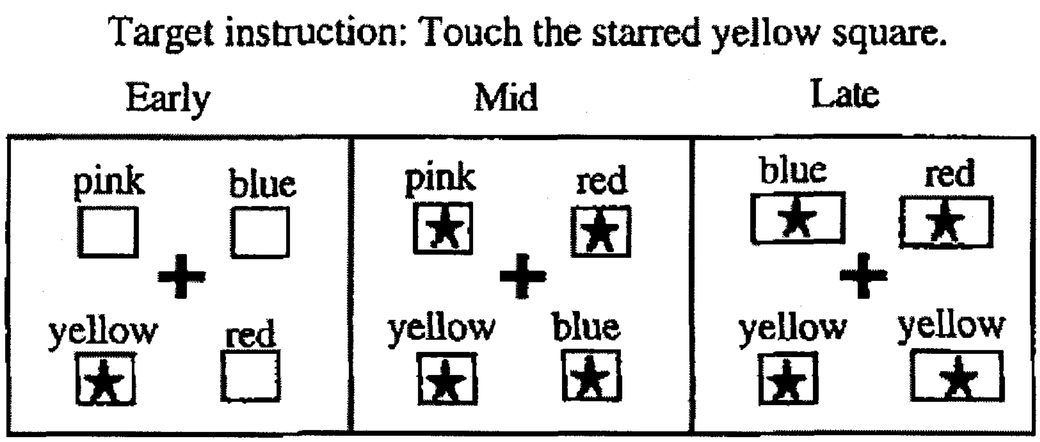


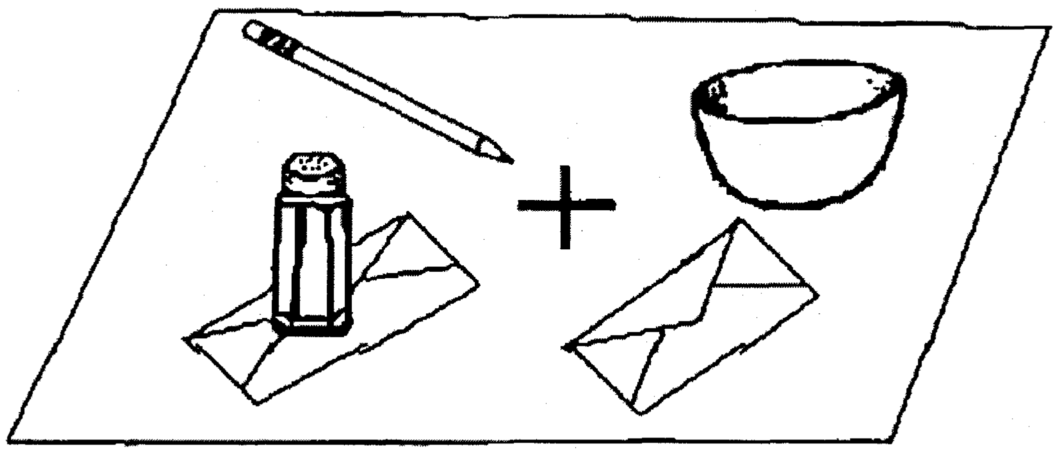
Figure 6. Example of the three point-of-disambiguation display conditions used in Tanenhaus et al. (1995) study (taken from Eberhard et al. 1995, p 415).

the sentence, and that they established reference as soon as distinguishing information was received. Participants typically initiated saccades within 300 ms of the onset of the disambiguating word (*starred*, *yellow*, and *square* respectively), therefore, before the onset of the noun (i.e., *square*) for the early and mid conditions. Tanenhaus and colleagues (Tanenhaus et al., 1995; Eberhard et al., 1995) argued that these results provided evidence that spoken language is processed incrementally and that visual context is rapidly integrated with spoken linguistic input.

In the final experiment, participants were presented with two sets of instructions: temporarily ambiguous instructions (e.g., "*Put the saltshaker on the envelope in the bowl*") and unambiguous control instructions (e.g., "*Put the saltshaker that's on the envelope in the bowl*"). In the first instruction, "*on the envelope*" is ambiguous because the listener does not know whether the speaker intends it to be the destination of the verb put (i.e., the place where "the saltshaker" is to be moved), or the modifier specifying the properties of "the saltshaker" to be moved (i.e., that is currently on "the envelope"). Similar to the reading study by Altmann and Steedman (1988), the authors provided participants with different visual contexts (rather than different discourse contexts), to examine whether syntactic processing of spoken input is initially structured independently of context (i.e., is encapsulated).

As shown in Figure 7, the visual context with which they were asked to interact consisted of two possible displays, one and two referent contexts. Tanenhaus et al., (1995) hypothesized that according to both the "incremental" and "encapsulated" models the one-referent context will support the destination interpretation. With the presence of only one saltshaker and the principles of minimal attachment and late closure, listeners

a)



b)

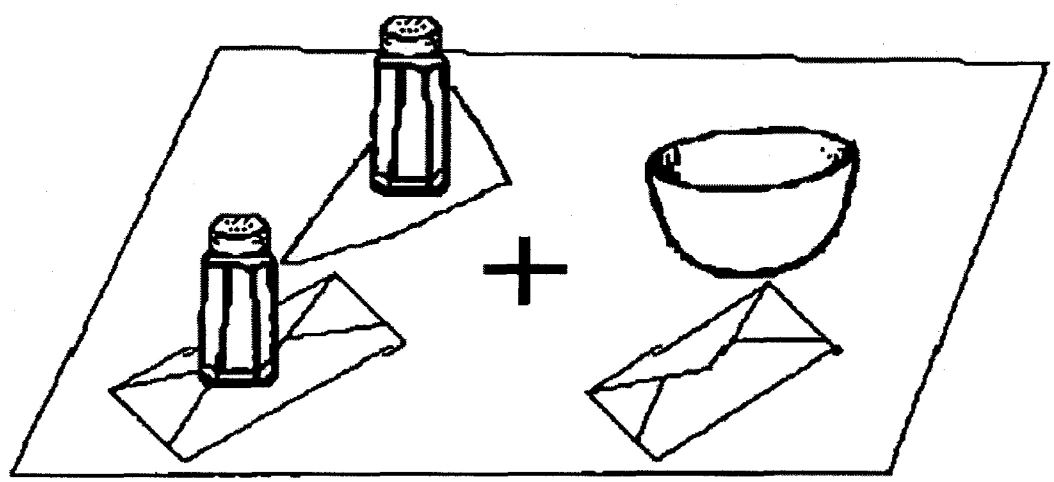
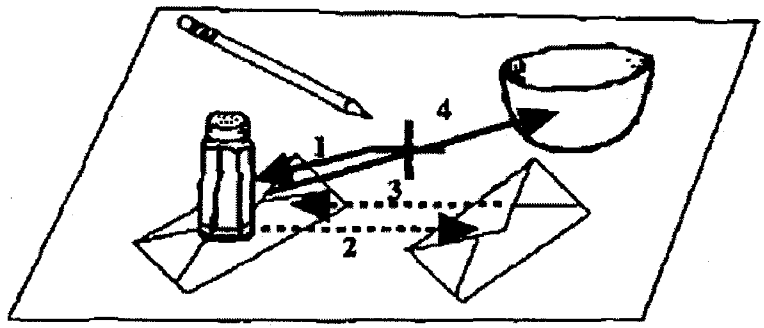


Figure 7. Example of the one (a) and two-referent (b) display conditions used in Tanenhaus et al. (1995) study (taken from Eberhard et al. 1995, p 427-428)

will assume that "*on the envelope*" refers to the destination of the verb "put." Realizing that they have been led the "garden-path" as they reach the word "*in*", a reanalysis will then be conducted in favor of the modification interpretation. On the other hand, according to the "incremental" processing model, they hypothesized that the two-referent context will support the modification interpretation. Because with the presence of two saltshakers the listener will correctly assume that "*on the envelope*" provides the necessary information for distinguishing which saltshaker is the intended theme. By contrast, according to the "encapsulated" models (e.g., Ferreira and Clifton, 1986; Frazier, 1987), listeners will once again interpret "*on the envelope*" as supporting the destination interpretation and the visual context will not prevent "garden-pathing."

As illustrated in Figure 8, participants made different sequences of eye-movements when the ambiguous instruction was given in the two types of visual context. In the one-referent condition, the sequence of eye-movements was as follows: (1) the target object 500 ms after hearing "*saltshaker*"; (2) incorrect goal referent 484 ms after hearing "*envelope*", thus, interpreting it as the destination (55% of trials); (3) a return to the target object after hearing "*bowl*"; and (4) the correct goal referent. On the other hand, with the unambiguous control instruction, participants never looked at the incorrect goal referent; rather they looked at the correct goal referent shortly after hearing "*bowl.*" In the two-referent condition, the sequence of eye-movements was the same for both the ambiguous and unambiguous instructions: (1) looked at both referents after hearing "*the saltshaker*"; (2) the target referent after hearing "*on the envelope*", thus the intended referent could not be identified until this point; (3) the correct goal after hearing "*the bowl.*" In the latter case, participants rarely (17% of trials) looked at the incorrect goal

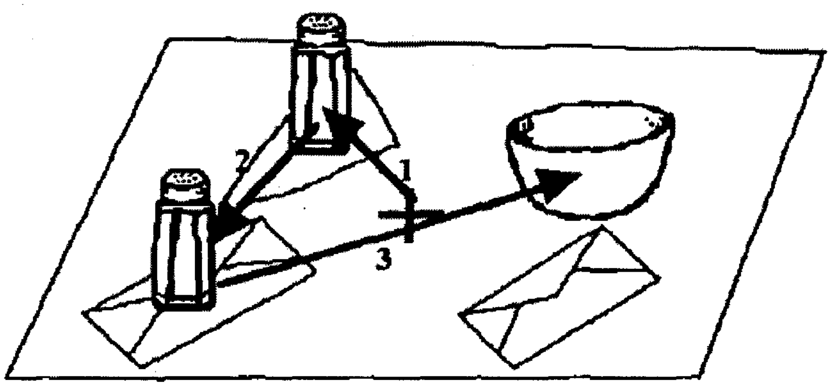
a)



Ambiguous: Put the saltshaker on the 1 envelope in 2 the bowl. 3 4

Unambiguous: Put the saltshaker that's 1 on the envelope in the bowl. 4

b)



Ambiguous: Put the saltshaker on the 1 envelope 2 in the bowl. 3

Unambiguous: Put the saltshaker that's 1 on the envelope 2 in the bowl. 3

Figure 8. Example of the sequence of eye movements to the objects in the one-referent (a) and two-referent display conditions, for both the ambiguous and unambiguous instructions used in Tanenhaus et al. (1995) study (taken from Eberhard et al. 1995, p 430)

referent (i.e., the envelope), suggesting that they correctly interpreted "*on the envelope*" as the modifier of the theme (i.e., the saltshaker). Eberhard et al. (1995) proposed that in situations in which listeners are directed to interact with objects in their environment, context becomes relevant to the comprehension process and affects the manner in which the linguistic input is initially structured.

In a similar series of studies, using the same methodology, Sedivy et al. (1999) examined how and when visual context mediates the interpretation of semantically ambiguous sentences containing scalar adjectives (e.g., tall, thin, etc.). Sedivy et al. (1999) assumed that scalar adjectives would be more difficult to interpret because they do not possess a stable core meaning, and rely strongly on the head noun or context to determine the value of its scale (e.g., *tall* in "tall building" is significantly larger in value than in "tall glass"). Participants were given instructions to move objects in a visual display (e.g., "Pick up the tall glass and put it below the pitcher"). As shown in Figure 9, the visual display contained a target object (e.g., a tall glass), a competitor object for which the scale was applicable (e.g., a pitcher), a distractor object (e.g., a key), and in half of the trials a contrast object (e.g., a short glass).

The results indicated that for displays which contained a contrast object, participants programmed their eye movements to the target object significantly sooner (sometime during the adjective), compared to the no contrast displays. Moreover, in the no contrast displays participants were significantly more likely to look at the competitor object, and the point of divergence between looks to the target object and competitor object occurred significantly later, than in the contrast displays. Finally, in the no contrast displays, looks to the competitor objects tended to occur earlier (at the word

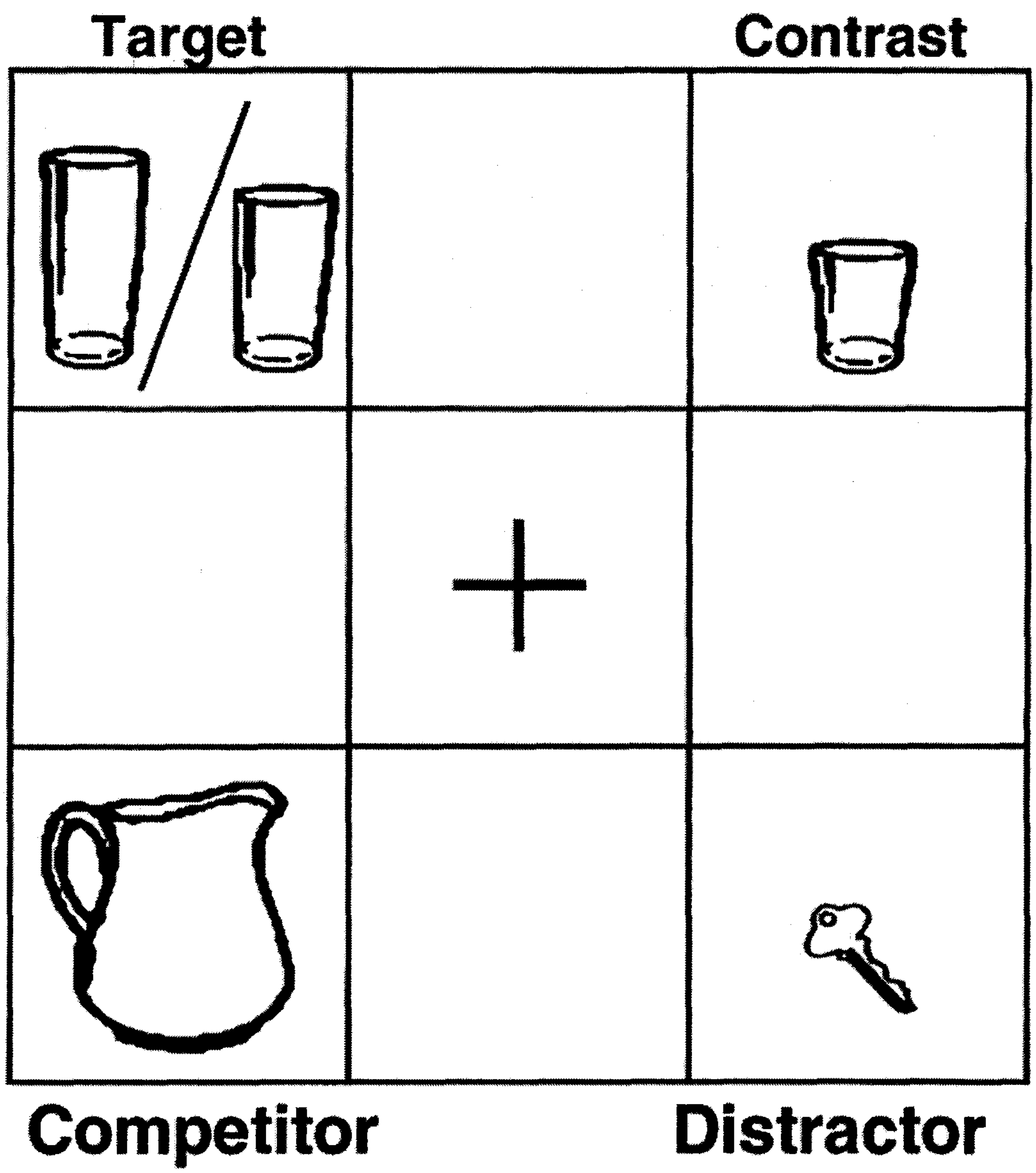


Figure 9. Example of a contrast display used in Sedivy et al. (1999) study (p. 130).

"*tall*") and drop off quickly as the noun head unfolded. On the other hand, looks to the contrast objects tended to occur later and not drop off as sharply, which Sedivy et al. (1999) interpreted as evidence of participants comparing the two candidates to ensure the correct target was chosen (i.e., the taller of the two glasses). Sedivy and colleagues (1999) concluded that the results suggest that semantic interpretation of adjectives occurs immediately and incrementally, even when they do not have a stable core meaning, by establishing contrasts between possible referents in the visual context.

Subsequently, in a series of studies Chambers et al. (2002) examined whether semantic-conceptual constraints extracted from spatial prepositions could immediately restrict visual attention to compatible noun phrase candidates. For instance, when hearing the instruction "Put the cube inside the can", the spatial preposition "*inside*" may limit attention only to referential candidates with container-like properties (e.g., a can or bowl). Moreover, they examined whether pragmatic considerations are also used when constraining the referential domain (e.g., not to any container, but to one that can accommodate the theme object). As shown in Figure 10, they used visual displays which contained two possible goal containers (e.g., a large and a small can), the theme object (e.g., cube), a distractor (e.g., bowl), and two unrelated objects. The size of the theme object was varied such that in some trials it could fit inside both goal containers and some trials only the larger container. Eye movement data indicated that participants limited their attention to goal objects that were compatible with the spatial prepositions and the theme objects. Chambers et al. (2002), thus, concluded that during the early moments of sentence processing the language system appears to integrate both linguistic and non-linguistic information in order to incrementally restrict the domain of referential

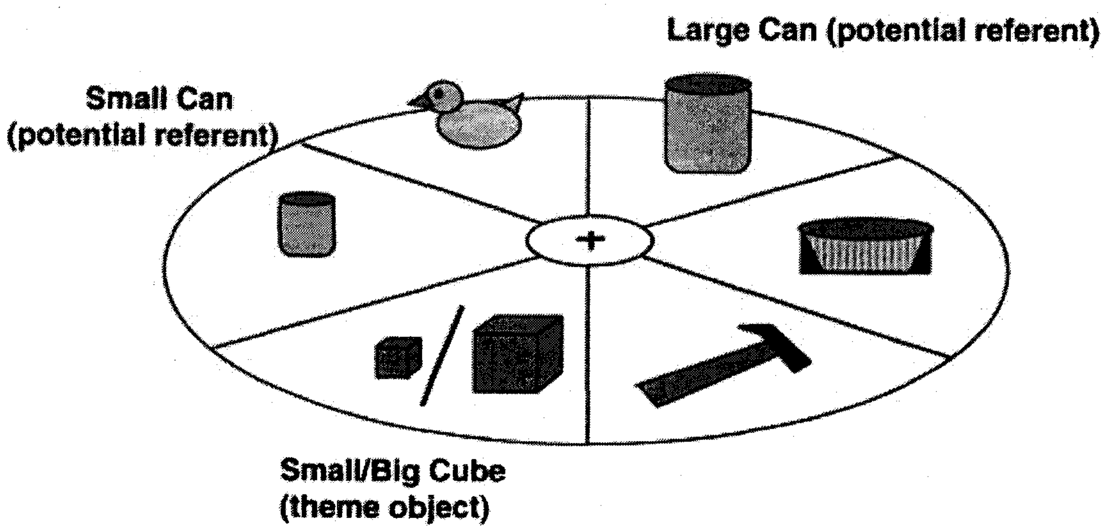


Figure 10. Example of a visual display used in Chambers et al. (2002) study (p. 38), with one (large cube) and two (small cube) compatible referent conditions.

candidates.

In sum, there is no disputing the fact that these studies demonstrate that the visual and linguistic systems interact at the earliest stages of processing. What is disputed, however, is how these results are then interpreted. Tanenhaus and his colleagues (Tanenhaus et al., 1995) argue that these results show that "referentially relevant nonlinguistic information immediately affects the manner in which the linguistic input is initially structured…[and] approaches to language comprehension that assign a central role to encapsulated linguistic subsystems are unlikely to prove fruitful." This conclusion is based primarily on the fact that the context-dependent ambiguity resolution occurred within approximately 250 ms after the offset of the critical linguistic stimulus. The eye-movement behavior may, however, reflect late, integrative processes rather than true contextual influences. Given that the visual system is taken to identify scenes in as fast as 100 ms (Potter, 1993), it is quite possible that the incoming linguistic information simply selected the appropriate representations of already activated visual referents in the scene, rather than being influenced by them.

Moreover, as mentioned by Tanenhaus and his colleagues (Tanenhaus et al., 2000), their set of instructions and visual displays may have encouraged participants to develop task-specific strategies that did not involve "normal" language processing. For example, when viewing the display portrayed in Figure 7, participants may have noticed that there were two saltshakers and, in turn, foresee that the upcoming instructions would refer to one of these two saltshakers and maybe even predict the form of the instruction. Empirical support for this last argument can be observed in participants' tendency to fixate the incorrect target object before fixating the correct one. This scan path is demonstrating that participants did not wait till the offset of the noun phrase before initiating their saccade, which may be because they felt pressured to quickly perform the task.

The Use of Verb-Specific Knowledge in Language Comprehension

Altmann and Kamide (1999) proposed that semantic information extracted from verbs could also guide visual attention towards the object referent of the verb's direct object, just as the adjectives and prepositions in the studies mentioned above served to narrow down the referential domain, *before* the actual onset of the noun. It is well known that verbs have different linguistic properties which determine the nature and types of internal and external arguments that combine to form larger units of syntactic and conceptual representations (see e.g., Grimshaw, 1990; Jackendoff, 1990). For instance, in the sentence offered by Mitchell (1989), "after the private (*saluted/fainted*) the sergeant he requested permission to end the exercise", the verb "saluted" would be grammatically correct, yet not "fainted." Because according to the syntactic frame in which each verb can occur (subcategorization features), "saluted" can be used either transitively (i.e., with a direct object – "the private saluted the *sergeant*") or intransitively (i.e., without a direct object – "the private saluted"), whereas "fainted" can only be used intransitively (e.g., "the private fainted" and not "the private fainted *the sergeant*"). In addition, the verb "saluted" has selectional rules (e.g., number, gender, case agreement) restricting the features of the subjects (e.g., +animate, +human, etc…) and direct objects (e.g., +animate, +human, +etc.) that can be associated with it. According to Chomsky (1965), breaking strict subcategorization rules leads to worse deviations from grammaticalness than breaking selectional rules (e.g., "the private fainted the sergeant" vs. "the private saluted the water").

Moreover, researchers agree that verbs play a key role in determining parsing preferences, either by constraining the nature of their syntactic complements (e.g., MacDonald, Pearlmutter & Seidenberg, 1994) or by triggering re-analyses of misparsed constituents within a sentence (e.g., Clifton, Frazier, & Connine, 1984). Tanenhaus, Carlson and Trueswell (1989) offer the example of the verb "*donate*", which they suggest gives a wealth of information, such as the meaning of the verb (e.g., to give), the thematic roles of the typical entities that will participate in the events described by the verb (e.g., requires an agent or donor, a patient or recipient, and a theme or the object that is donated), guide the assignment of noun phrases (i.e., referents of real-world people and objects) to the various roles licensed by the verb; and the argument structure of the verb (requires at least two argument Noun Phrases (NP), and an optional NP). Moreover, it has been theorized (see MacDonald et al., 1994) that not only does information extracted at verbs serve to identify the roles associated with the action, but also the position within the sentence where the recipients of those roles can be found (e.g., English is a subject-verb-object language). For instance, the sentence: "*John (subject NP/agent) donated fifty dollars (object NP/theme) to his favorite charity (PP/recipient)*" satisfies the abovementioned criteria.

As mentioned above, Altmann and Kamide (1999) set out to examine whether semantic information extracted from verbs can guide visual attention towards the object referent of the verb's direct object, *before* the noun is uttered within the sentence. Participants were presented with semi-realistic visual scenes made up of color drawings and one of two possible spoken sentences. For example, in the visual scene depicted in

Figure 11 there is a young boy sitting on the floor surrounded by various items (toy train, toy car, birthday cake and a ball). One of the sentences (e.g., *"The boy will eat the cake"*) contained a verb (i.e., "eat") that was semantically restrictive (i.e., its direct object had to be something edible), which allowed only one object (i.e., "cake") in the visual scene to be the post-verbal target. By contrast, the other sentence (e.g., "*The boy will move the cake*") contained a verb (i.e., "move") that was less semantically restrictive (i.e., its direct object has to be something moveable), that allowed any object within the visual scene to be the post-verbal target object. Filler items were added in which the target object referred to in the sentence was not present, in order to allow for a judgment task at the end of each trial – respond "yes" or "no" whether the sentence could apply to the scene presented.

Altmann and Kamide (1999) measured proportion of saccades launched towards the visual referent of the verb's direct object (e.g., "cake") before the onset of the noun, and saccade onsets towards the target object relative to the onset of the verb. Eye movement data indicated that the target object was fixated in 90% of the trials, and that on 54% of the semantically restrictive trials (e.g., "eat cake") first saccades to the target object were launched prior to the onset of the noun, as compared to 38% of the semantically non-restrictive trials (e.g., "move cake"), yet this was not statistically significant. Moreover, they found that the first saccade to the target object was launched 85 ms *before* the onset of the post-verb noun in the semantically restrictive condition, and 127 ms *after* the onset of the noun in the semantically non-restrictive condition. Altmann and Kamide (1999) concluded that these results suggest that the activation of verb-specific information can direct eye movements towards objects in the visual scene that

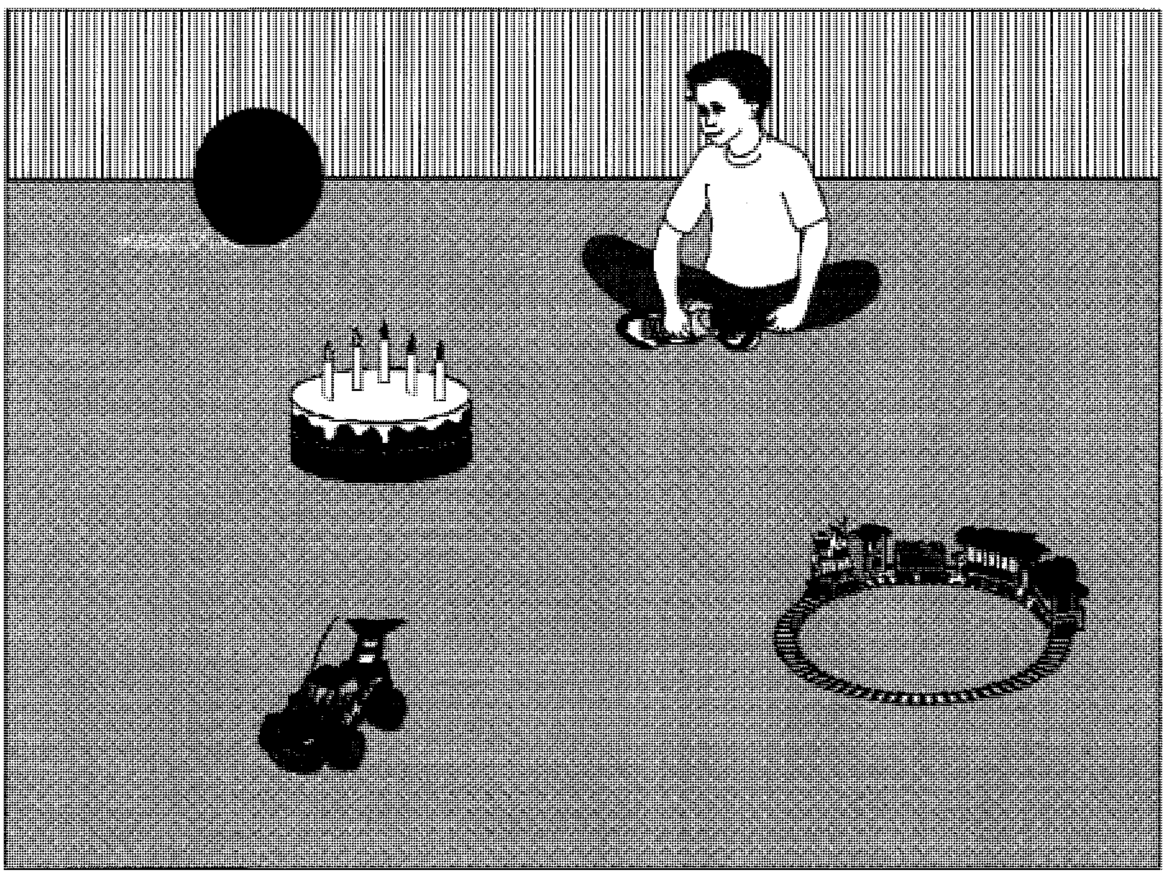


Figure 11. Example of a scene used in Altmann and Kamide's (1999) study (p. 250). The accompanying sentence was either "The boy will *eat* the cake" or "The boy will *move* the cake."

are consistent with the selectional restrictions of the verb, even before the semantic properties of the direct object become available.

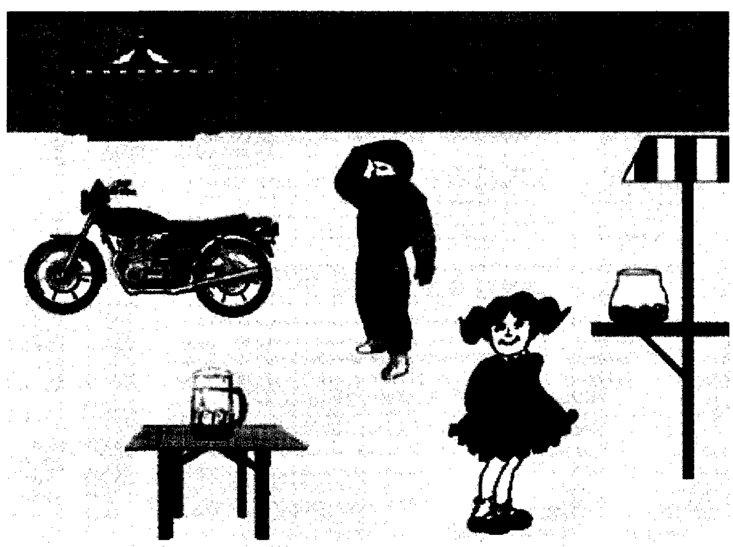
Altmann and Kamide (1999) carried out a second experiment identical to the first, but without the judgment task. It was hypothesized that task demands may have prompted participants to develop a strategy which induced anticipatory eye movements not normally seen in everyday language processing. The pattern of eye movements were similar to experiment 1; with significantly more fixations to the target object before noun onset for the restrictive trials than the non-restrictive trials (32% and 18% respectively), and significantly faster saccade onset times in the restrictive condition than the non-restrictive condition (291 ms and 536 ms *after* noun onset, respectively). Altmann and Kamide (1999) argued that although the saccade onset times were much longer (approximately 350 ms) in the absence of a decision task, the eye movement data still indicated that verb-specific information can be used to guide attention towards the appropriate visual referent of the verb's direct object. Altmann and Kamide (1999) thus concluded that verbs contain syntactic/semantic information that can be used by the language processor to restrict and in some cases predict the identity of the intended visual referent before it is encountered in the linguistic input.

Using the same methodology, in a series of studies Kamide, Altmann and Haywood (2003) examined whether verb-specific information could also guide eye movements towards the second post-verbal argument (i.e., the goal) prior to its onset within the sentence. The data for this study (see Figure 12a) demonstrated that during the theme expression ("the butter") there were more anticipatory eye movements towards the appropriate (i.e., fulfilling the semantic restrictions of the verb) goal object ("the bread"

a)

The woman will	spread	the butter	on the bread.
The woman will	slide	the butter	to the man.

b)



The man will	ride	the	motorbike.
The girl will	ride	the	carousel.
The man will	taste	the	beer.
The girl will	taste	the	sweets.

Figure 12. Example of scenes used in Kamide et al. (2003) study (p. 138 and 140).

in the case of the verb "spread" or "the man" in the case of the verb "slide") than towards the inappropriate goal object ("the man" or "the bread" respectively). It was unclear from these data, however, whether it was the verb alone ("spread" or "slide") that resulted in the anticipatory eye movements to the correct goal object, or a combination of the verb and its direct object ("spread the butter" or "slide the butter").

These results led Kamide et al. (2003) to examine whether information about the agent (i.e., entity that instigates the action) combined with the selectional restrictions of the verb to predict the theme (grammatical patient or direct object). The data (see Figure 12b) suggested that the anticipatory eye movements to the verb's direct object were the result of combinatory effects. As there were increased anticipatory looks to the "motorbike" in "the man will ride" condition than in "the girl will ride" condition, even though both themes satisfied the selectional restrictions of the verb "ride." Moreover, the anticipatory eye movements were not the result of direct associations between the agent and the theme (man and motorbike), independent of the verb "ride", as there were no increased looks to the "motorbike" in "the man will taste" condition than in "the girl will taste" condition. Taken together, these results further support Altmann and Kamide's view that the language processor incrementally uses all available information (linguistic and non-linguistic), such as the selectional restrictions afforded by the agent and verb, to guide visual attention towards the appropriate goal or theme object, prior to its utterance within the linguistic input.

In sum, most studies cited above have focused on the influence of visual context in lexical and sentence processing, from syntactic to semantic computations. The results all seem to point in the same direction: visual context plays a key role in the analysis of linguistic information, and that language processing is not at any degree autonomous in its processing routines, rather is influenced by the demands of (visual) context. Tanenhaus, Altmann and their colleagues base this conclusion on the observation that context-dependent ambiguity resolution occurs within a short time after the critical linguistic stimulus, and that anticipatory eye movements to the verb's direct object can occur prior to its utterance within the linguistic input. Yet, once again, the fact that visual context can quickly interact with language processing, does not necessarily point to the inability of the perceptual systems to produce context-independent representations (i.e., autonomous processing). As discussed above, it is possible that initially the linguistic module phonologically decodes, structures, and syntactically parses the incoming linguistic information, and that this output is then matched against the conceptual representations activated by the visual context.

The locus of the interaction between modules may be within what Potter (1976, 1993, 1999) refers to as conceptual short term memory (CSTM). CSTM is a form of short-term memory where conceptual representations are rapidly activated and held momentarily during the early stages of perceptual processing, memory retrieval and thought (Potter 1993). Potter (1976, 1993, 1999) suggests that when stimuli such as objects are recognized, their conceptual information becomes quickly activated. These activated representations are then linked to semantically related information stored within long-term memory (LTM), and thereby activates this information. According to this model, the concurrently activated representations are then conceptually structured and finally consolidated into LTM. Given the large number of activated representations at any one time, only information that can be organized into a meaningful structure is integrated within LTM, while the rest is quickly forgotten. As some properties of the CSTM satisfy Fodor's (1983) module criteria (i.e., quick, unconscious, and not under conscious control), whereas others satisfy his central level processor criteria (i.e., interaction between modules and access to information in long-term memory), one possibility is that the CSTM is a post-modular but pre-central process. The studies in this paper will, therefore, further explore the nature of the interaction between the language and visual systems. The results will then be discussed in the context of a proposed model, inspired by Potter (1993, 1999), in which the CSTM is responsible for the interaction and integration of activated conceptual representations outputted by each module.

Rationale and Purpose of this Study

The first set of experiments was designed to investigate the process of interaction between language and vision using a cross-modal eye-tracking method. This methodology was used to trace the time-course of integration between linguistic representations and representations computed by the visual and conceptual systems. Crucial to the dispute about whether or not the language processing system is influenced by the immediate context is an understanding of how and when different types of verb-specific representations are accessed. Of particular interest was understanding how selectional restrictions, conceived as the different types of noun complements that verbs license, direct the attention system (which, by hypothesis, controls eye-movements) in search of referents of their grammatical objects at different points during sentence comprehension. This study addressed two major questions regarding the use of verbs in language comprehension: (1) what types of verb-specific information are computed during sentence comprehension and (2) what is the time course of access to such information. Underlying these two general questions is the investigation of how visual representations may influence the selection of conceptually plausible grammatical objects of verbs.

Analogous to Altmann and Kamide (1999), the role of verb-specific information in guiding visual attention towards the object referents of the verb's direct object was studied. Of particular interest was a subcategory of result verbs called causative verbs (also known as accomplishment verbs), so labeled because they specify an action and the result of that action (e.g., the verb "burn" causes an object to become burned). In the visual world paradigm this class of verbs is of practical use for two important reasons: (1) the actions denoted by the verbs can be represented visually, or alluded to, using static displays (e.g., a woman with a lit match next to a candle); and (2) they are semantically restrictive, in that the direct objects they can receive are constrained by the semantic properties of the verb (e.g., burnable). For example, the verb "burn" can only take patients that are burnable, such as candles or other flammable substances, and not for instance an iron substance which does not possess the physical/chemical property of being burnable. In our study the causative verbs were contrasted against perception and psychological verbs, so named because they specify the perception of an object (e.g., the verb "felt") or a psychological event (e.g., the verb "examined") (see Levin, 1993, for a classification of verb classes). These verbs are considered less semantically restrictive, as compared to the causative verbs, in that they can refer to any direct object that is perceptible.

Similar to the results obtained by Altmann and Kamide (1999) and by Tanenhaus and colleagues (e.g., Eberhart et al., 1995), it was assumed that there would be "anticipatory" eye movements (as revealed by faster saccade onset times (SOT)) and less first saccade errors in the more semantically restrictive condition. Furthermore, as in their study, filler items were added, such that the target object referred to in the sentence was not present, in order to allow for a "yes/no" judgment task at the end of each trial. This set of studies, however, differed significantly from Altmann and Kamide's (1999) study by attempting to correct for some important methodological problems, as discussed in the following paragraphs.

First, Altmann and Kamide's (1995) participants were free to move their eyes prior to verb onset. Thus participant's eye gaze position, relative to the target object, at verb onset was not held constant across trials. When using time to initiate a saccade to the target object relative to verb onset as a dependent measure, it is important to control for starting eye gaze position, because as observed by Altmann and Kamide, in some trials participants were already fixating the target object at verb onset, which potentially resulted in an overall underestimation of SOT values. Furthermore, the objects in each scene were haphazardly placed (e.g., a lit cake located on the floor) and, therefore, not equidistant from each other. For some trials, this confound may have resulted in quicker saccade onset times simply because the target object was closer to the current gaze position. To prevent this occurrence, saccade onsets and fixations were controlled in this study by presenting a red dot at the center of the screen, equidistant from a set of six pictures of objects. Participants were asked to fixate this red dot until it turned green, which alerted them that they were now free to move their eyes. The onset of the green dot occurred at 6 different points during the presentation of the sentence; 3 within the verb (onset, middle and offset), and 3 within and after the noun (onset, offset, and offset+200 ms).

Second, Altmann and Kamide did not collect normative data for their conceptually related noun-verb pairs. Tesolin and de Almeida (2000) collected such norms by asking participants to fill in sentence frames with the most appropriate noun (e.g., "The person burned the ____" or "The ____ burned"). In order to eliminate close associates of those verbs – on the assumption that the procedure of asking for the "first thing that came to mind" could have elicited conceptually-related but also close associations of the verbs, additional participants were administered the list of verbs without sentence frames (e.g., "burn____"). If the strong verb associates matched the most appropriate noun they were eliminated from the set. This created a final set of 24 syntactically and conceptually related verb-noun pairs (e.g., "burn-candle"). Because Altmann and Kamide did not collect these normative data, their results could be explained by stronger verb associations in their semantically restrictive conditions (e.g., "eat-cake"), than their less semantically restrictive conditions (e.g., "move-cake").

Third, Altmann and Kamide did not collect normative data for the pictures they chose to depict their target objects. Thus, it is not certain whether the names that the participants were assigning to each picture were the ones intended by the experimenters. This possible confound was eliminated by conducting a normative study in which pictures of the target objects were flashed on a screen and participants were asked to name the objects depicted. Moreover, the pictures used in their studies were line drawings, whereas colored photographs of objects were used in this study, thus improving ecological validity.

Fourth, Altmann and Kamide used semi-realistic scenes, in which some contained helpful additional cues, as in the example provided in Figure 11. The agent in this scene (i.e., the boy) is looking directly at the target object (i.e., the cake). This confound also occurred in two other scenes. Thus, even before the presentation of the verb participants could deduce from the agent's stare the identity of the upcoming target. In order to avoid providing participants with unwanted additional cues, the pictures of objects were presented in a circular pattern, with no picture representing the agent in the sentence.

Fifth, in addition to the picture of the verb's direct object, the 24 sets of pictures also contained linguistic (phonological), conceptual (semantically related to the target), and visual (shape and color) competitors. Through the use of these competitors, it was possible to examine the time course of activated representations (linguistic, conceptual, or visual) as information about lexical alternatives started to become available.

Sixth, the duration of pre-exposure to the objects in the visual scene was manipulated (6 and 24 seconds) to examine whether increased scanning time would produce more accurate and stable memory representations, as indicated by quicker SOTs to the target objects and fewer first saccade errors.

Seventh, Altmann and Kamide did not restrict participant's head movements to assure that the distance and, therefore, visual angle between the computer screen and the participant's eyes remained constant across trials and participants. In this study, head movements were restricted by having participants place their head on a chin rest 57 cm from the computer screen. This restraint also guaranteed that throughout the experiment participants were indeed facing the visual display.

Experiment 1. Monitoring Eye Movements during Sentence Comprehension using Static Displays

In order to investigate the interplay between verb representations and the visual context, four normative studies and three experimental studies were conducted. In the normative studies, data were collected on verb-complement information (Study 1), picture naming (Study 2), name-picture matching (Study 3), and object features (Study 4). In Experiments 1a, 1b and 1c the eye movements to objects were tracked during the presentation of sentences. Data were collected at six probe points (points at which the onset of eye-movement recordings were done), three at the noun direct object of the main verb (Experiment 1a) and three at the main verb, with 6 (Experiment 1b) or 24 seconds (Experiment 1c) to encode the visual display.

1.1 Normative Studies

1.1.1 Verb-Complement Selection

To select the appropriate preferred direct object noun for a set of lexical causative verbs, Tesolin and de Almeida (2000) collected verb-complement information using a paper-and-pencil task. Causative verbs were used because they impose strong selectional restrictions on the nature of the object of the causative structure (the object that, by hypothesis, undergoes a change of state). Twenty-four participants were asked to fill in the blanks with the first word that came to mind that fit the context of both sentence frames (e.g., "*The person bounced the_______*" and "*The_______ bounced*"). In order to eliminate close associates of those verbs, another 10 participants were asked to write down the first word that came to mind for each verb without the sentence frame (e.g., *bounce* _______) (see Tesolin & de Almeida (2000) for additional information regarding methodology and results). From the initial 47 verbs, 24 were selected based on the following two criteria: (1) the lexical causative verbs had strongly preferred complements (e.g., *ball* for *bounce*), and (2) the complement referents could be pictured by concrete basic-level objects, rather than abstract hard-to-picture objects (e.g., *snow*) or superordinate categories (e.g., *food*). When the most frequent noun provided for each verb did not meet these criteria, the second or third most frequent noun was selected (e.g., *grapes* instead of *ice* for the verb *crushed*).

Once the main set of verb-noun pairs was selected, pictures for each noun referent were chosen. Moreover, five pictures of objects were selected that resembled each verb-complement object in (a) initial phonological segment (e.g., a candy for *candle*), (b) meaning or function (e.g., a lamp for *candle*), (c) visible shape (e.g., cylindrical and elongated), (d) color (e.g., white and gold), and (e) saliency, but with neither of the former characteristics (i.e., unrelated). These sets of objects are hereafter referred to as "competitors", because as will be explained in detail below, they are believed to compete with the noun referent for attentional resources. All these pictures of objects were selected from Hemera's Photo Objects Premium Image Collection CDs (version 2.02) and scaled to fit 150 pixels in the horizontal and/or the vertical direction, depending on the overall shape of the object, over a 21″ (1024 x 768 pixels) screen monitor.

1.1.2 Picture Naming

To ensure that the sets of selected objects were recognized equally well across categories (i.e., that candle, candy, lamp, etc., had similar naming latencies), and that the names assumed to be the main labels for each object were correct (e.g., that *candle* was not going to be called by any other name), a picture naming task was conducted (Van de

Velde, 2001). In this task participants were asked to name each object as it appeared on a computer screen. The data demonstrated that participants on average correctly identified 83% of the pictures, and took on average 1080 ms to name each picture (see Van de Velde, 2001, for additional information regarding methodology and results).

1.1.3 Name-Picture Matching

Since the goal of the main experiments was to match the pictures of objects with the noun's referent, a name-picture matching task was conducted for the items that produced less than 60% accuracy in naming or that were not named within the allotted 2500 ms presentation time (Van de Velde, 2001). In this task participants had to indicate which of the three pictures of objects on the screen corresponded to the word (i.e., name of an object) they just heard. Participants responded correctly to all the target pictures of objects. The pictures that produced the shortest reaction times (RT) were then chosen for the purposes of the main study. Of the original 36 pictures in the picture naming study that were problematic: 20 were kept, 11 were replaced by a more suitable alternative (e.g., a more easily recognizable picture of a *candy*), and 5 were replaced by completely new pictures of objects (for the unrelated competitor category only) (see Van de Velde, 2001, for additional information regarding methodology and results).

1.2 Experiment 1a: Cross-modal eye-tracking with sentences – probing at the noun

The goal of this set of experiments was to assess the effect of both linguistic and visual contexts on the pattern of eye movements as participants listened to sentences and looked for sentence-related objects on a screen. More specifically, the semantic-selection properties of different classes of verbs and how these may be further constrained by information provided by the visual context was examined. Two different types of sentence contexts were used, one with a semantically restrictive causative verb (e.g., *The waitress burned the candle*) and another with a less-restrictive perception verb (e.g., *The waitress admired the candle*). Eye movements were monitored at three points during the presentation of the noun that served as direct object of the main verbs in the sentences (e.g., *candle*): at noun onset, noun offset, and 200 ms after noun offset. The two classes of verbs were expected to produce different patterns of eye movements. According to the results obtained by Tanenhaus, Altmann and their colleagues (e.g., Eberhart et al., 1995; Altmann & Kamide, 1999), it was hypothesized that the more restrictive causative verb class would result in "anticipatory" eye movements to the target object (as revealed by shorter SOT), as well as fewer first-saccade errors, as compared to the less restrictive perception verb class. If it were assumed, however, that the linguistic processor operates independently of the information provided by the visual context, one would not expect to find any difference in the processing of the noun target across the different verb conditions.

Because the three experiments had similar methodologies, the materials, design, and procedure used for the first experiment (Experiment 1a) are described in detail, followed by the modifications made to Experiments 1b and 1c. The results section, which comprises the analyses conducted on all three experiments is collapsed into one comprehensive section and presented after the three methodology sections. This type of layout was chosen to allow for greater ease in comparing participants' performance across all three experiments.

1.2.1 Method

1.2.2 Participants

Thirty Concordia psychology undergraduate students participated in this experiment for course credit or monetary compensation. All were native English speakers and had normal or corrected-to-normal vision. None of the participants took part in the normative studies.

1.2.3 Material and Design

The stimuli for this experiment consisted of the sets of object pictures selected in the normative studies described above. There were 144 pictures of objects divided into 24 sets of 6 objects (see Appendix A2 for the set of pictures). The target picture (e.g., *candle*) in each set represented the grammatical direct object of the verb in the sentences selected in the normative study described above. The other pictures represented a phonological competitor (e.g., *candy*), a semantic competitor (e.g., *lamp*), a shape competitor (e.g., *pen*), a color competitor (e.g., bottle of perfume) and an unrelated competitor (e.g., cake).

The sentences (see Appendix A2) were spoken by a female student and recorded at a natural pace on an Apple Macintosh using SoundEdit at 16 bits and 44.1 kHz. The sentences consisted of two different types of sentential contexts, one with a semantically restrictive causative verb (e.g., "*The waitress burned the candle*") and another with a less-restrictive perception verb (e.g., "*The waitress admired the candle*"). All main clauses were of the form NP1(Noun Phrase1)-*will*-Verb-NP2. The NP1 always made reference to the agent in generic form (*the boy, the painter, the woman*, etc.) and the NP2 made reference to the target object in the display. Each participant was exposed to all six experimental conditions (Noun Onset-Causative, Noun Offset-Causative, Noun Offset+200ms-Causative, Noun Onset-Perception, Noun Offset-Perception, Noun

Offset+200ms-Perception), but only one condition for each scene. Out of a 144 possible stimulus combinations (three probe points X two verb conditions X 24 scenes), each participant was randomly assigned to one of six groups, each containing 24 items from the six conditions. The experiment, therefore, consisted of 60 trials: 12 trials with causative verbs and the target object present, 12 trials with perceptual verbs and the target object present, 24 trials with the direct object not present, six trials with random onset of pictures and the target object present, and six trials with random onset and the target object not present. These last 36 fillers (see Appendix A3) were to prevent participants from predicting the correct response (i.e., whether the target object was present in the visual display) and the onset of the pictures of objects.

The experiment was conducted on a Macintosh G4 computer with a Sony Trinitron Multiscan E500 21" color monitor (75 Hz refresh rate), placed 57 cm in front of the participant. The visual and auditory stimuli were presented via PsyScope (Cohen, MacWhinney, Flatt, & Provost, 1993). Eye movements were recorded using an SR Research EyeLink I head-mounted eye tracker at a sampling rate of 250 Hz. The eye tracker was controlled by a Pentium-III PC, which received the eye movement data (saccades, fixations, and blinks), and was coupled with a Macintosh computer via an Ethernet connection. Participants were seated in a chair, had the eye tracker adjusted on their heads, and their head movements minimized with the use of a chin rest. Minor head movements were corrected by a system of four LED sensors affixed to each corner of the monitor, and a head-mounted camera that continuously measured participants' head position in relation to the screen. The eye tracker camera was placed in front of the left eye (no data were collected from the right eye). Participants also wore a pair of headphones that wrapped around the neck so as not to interfere with the head-mounted eye-tracker.

1.2.4 Procedure

After completing the consent form (see Appendix B2), participants received instructions on the screen (see Appendix C1), which were then reinforced by the experimenter. For each participant, the eye-tracker was then manually calibrated to ensure that the device was properly adjusted to accurately record the eye movements across the entire screen. Following the practice trials, as a reminder of the task at hand, participants received a shorter version of the instructions (see Appendix C2). Figure 13 illustrates the sequence of events for each trial: (1) an asterisk appeared in the center of the screen, which served as a fixation point and preparation for the trial; participants were instructed to press a yellow button on a Carnegie Mellon University (CMU) response box to initiate the trial; (2) a set of six pictures of objects appeared and the asterisk was replaced by a green dot; participants were instructed that green indicated that they were free to move their eyes across the screen; the pictures and the green dot remained on the screen for 6 seconds; (3) the pictures then disappeared and the green dot was replaced by a red dot; participants were instructed that the red dot indicated that they were required to fixate back on the center of the screen and could not move their eyes until the dot again turned green; (4) the sentence was presented binaurally through headphones; and (5) simultaneously the dot turned green and the pictures reappeared in their original locations, at three different points during the presentation of the noun: noun onset, noun offset, and 200 ms after the onset of the noun. This prevented participants from having to inhibit a saccade or launching a saccade or using visual attention prior to the onset of

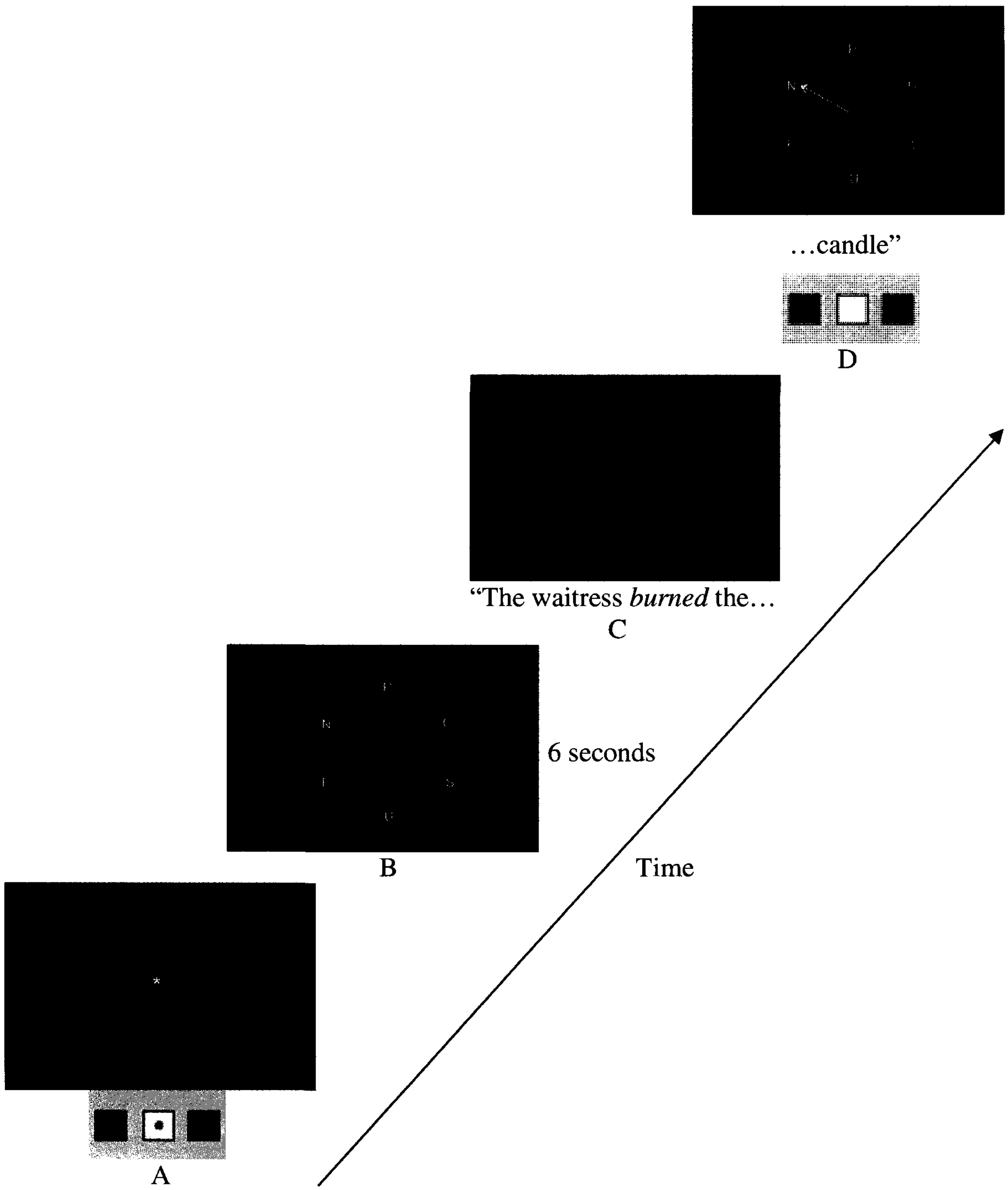


Figure 13. Time sequence of events in Experiment 1a: A) Asterisk was fixated until participants pressed the yellow button to begin; B) 6 pictures of objects appeared and center dot was green indicating that participants could freely scan the pictures for 6 seconds; C) pictures disappeared, center dot turned red (which they were asked to fixate), then sentence began; and D) at one of 3 probe points (noun onset, noun offset or noun+200 ms) 6 pictures of objects re-appeared in the same order, center dot turned green, and if picture of object mentioned in sentence was present, participants were asked to quickly fixate it and press the green button, if not they were asked to press the red button.

the green dot. Participants were asked to make a yes/no decision about whether the object they just heard in the sentences was present. If the object was present, participants fixated the object and then pressed the green button; if it was not present they pressed the red button on the CMU response box. The center red and green dots had a radius of 13.5 pixels. The pictures had the same dimensions as those used in the normative studies described above. The six pictures were presented in a circular arrangement: One picture was presented above fixation at 0°, below fixation at 180 °, two pictures were to the right of fixation at 60 ° and 120 °, and two were to the left of fixation at 240 ° and 300 °. The target object was presented in random positions across trials. The experiment lasted approximately 30 minutes (including eye-tracker calibration).

1.3 Experiment 1b: Cross-modal eye-tracking with sentences – probing at the verb

An additional study was conducted using the same materials, design and procedures as in Experiment 1a, with the exception that the probe occurred during the presentation of the verb (verb onset, middle of verb, and verb offset), rather than the noun. This study was conducted to examine whether: 1) the mediating effects of restrictive verb types on eye-movements occurs at earlier stages of sentence processing, well before the onset of the post-verbal noun, and/or 2) preventing participants from making eye movements until after noun onset may have suppressed any anticipatory eye movements.

1.3.1 Method

1.3.2 Participants

Thirty Concordia psychology undergraduate students participated in this experiment for course credit or monetary compensation. However, because of an eye-tracker malfunction one subject had to be removed from the analysis. All were native English speakers and had normal or corrected-to-normal vision. None of the participants took part in the previous experiment or normative studies.

1.3.3 Materials and Design

The same materials and design used in Experiment 1a were employed in this experiment.

1.3.4 *Procedure*

The procedure was the same as the one adopted in Experiment 1a, with the exception that the onset of the green light (the time at which the fixation point turned from red to green, thus allowing participants to make an eye movement) occurred at three different points during the presentation of the verb: verb onset, middle of the verb and verb offset (see Figure 14).

1.4 Experiment 1c: Cross-modal eye-tracking with sentences – probing at the verb with a longer inter-stimulus interval (ISI)

Six seconds may not have provided participants with sufficient time to identify and process the location of 6 pictures of objects (i.e., 1 second per object). Therefore, for this study, initial scanning time was increased to 24 seconds (i.e., 4 seconds per object). The expectation was that additional scanning time would lead to a more accurate internal representation of the visual display and therefore reduce first saccade errors to the target objects.

1.4.1 Method

1.4.2 Participants

18 Concordia psychology undergraduate and graduate students participated in this

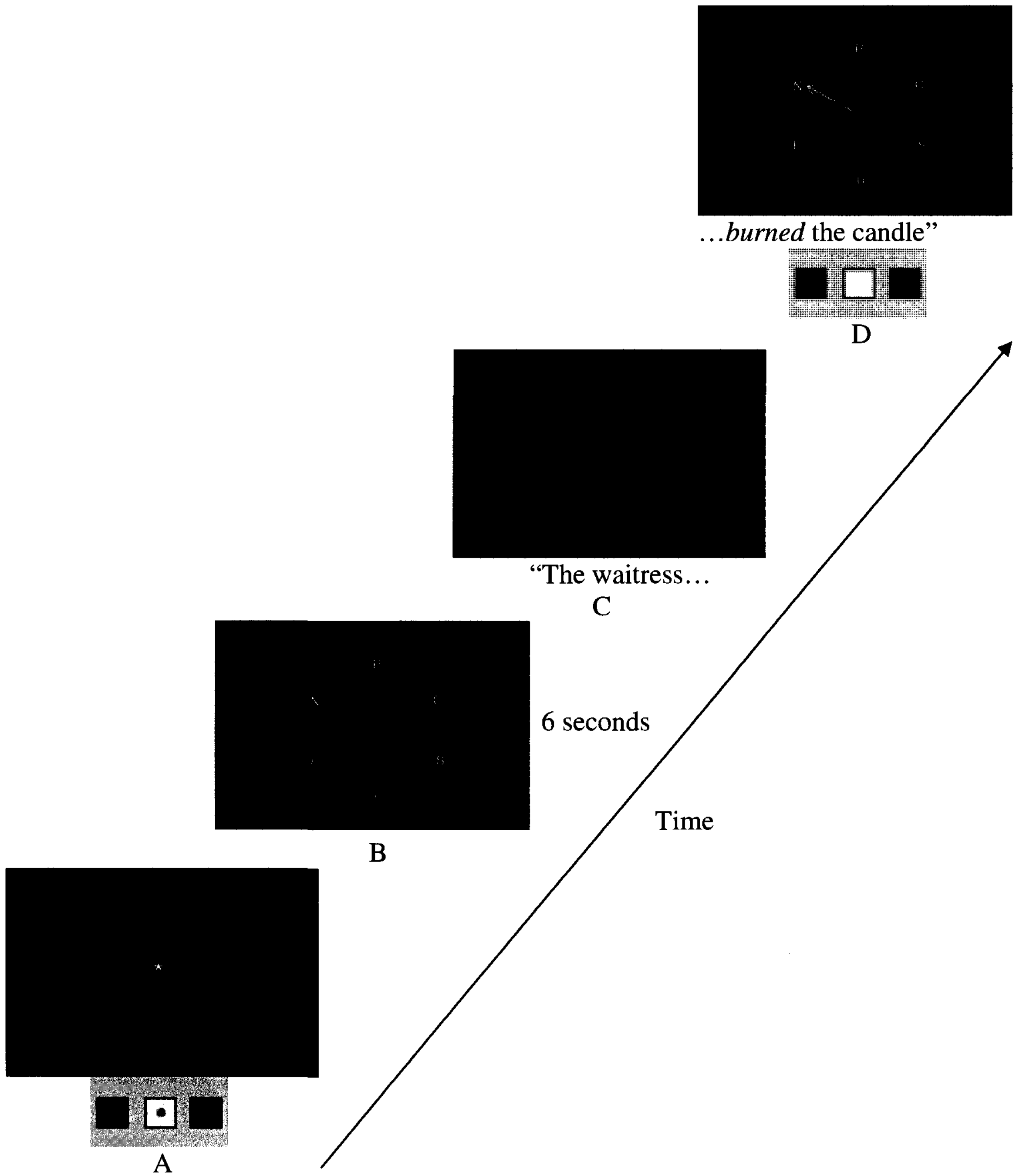


Figure 14. Time sequence of events in Experiment 1b: A) Asterisk fixated until participants pressed the yellow button to begin; B) 6 pictures of objects appeared and center dot was green indicating that participants could freely scan the pictures for 6 seconds; C) pictures disappeared, center dot turned red (which they were asked to fixate), then sentence began; and D) at one of 3 probe points (verb onset, middle of verb or verb offset) 6 pictures of objects re-appeared in the same order, center dot turned green, and if picture of object mentioned in sentence was present, participants were asked to quickly fixate it and press the green button, if not they were asked to press the red button.

experiment for course credit or monetary compensation. All were native English speakers and had normal or corrected-to-normal vision. None of the participants took part in the previous experiment or normative studies.

1.4.3 Materials and Design

The same materials and design used in Experiment 1a and 1b were employed in this experiment.

1.4.4 Procedure

The procedure was the same as the one adopted in Experiment 1b, with the exception that the participants had 24 seconds (rather than 6 seconds) to view the initial set of 6 pictures of objects (see Figure 15).

1.5 Results

The EyeLink Data Viewer software was used to produce a scan path of the saccades and fixations across the scene. Trials were excluded if eye fixations were not at the fixation point at the onset of the green dot or if no eye movements were made before responding on the button box (13.89%; 11.9%; and 6.71% of the trials for Experiment 1a, 1b, and 1c, respectively). To obtain saccade onset times (SOT), the main variable of interest in these studies, the difference between the onset of the green dot and the time to initiate a saccade to an object was calculated. Unless otherwise specified, data were analyzed using two two-way repeated measures ANOVAs. One analysis treated participants as a random effect and verb types and probe points as within-subject factors [*F1*], whereas the other treated items as a random effect and verb types and probe points as a within-item factors [*F2*]. Several sets of analyses were conducted on the eye movement data, as described below.

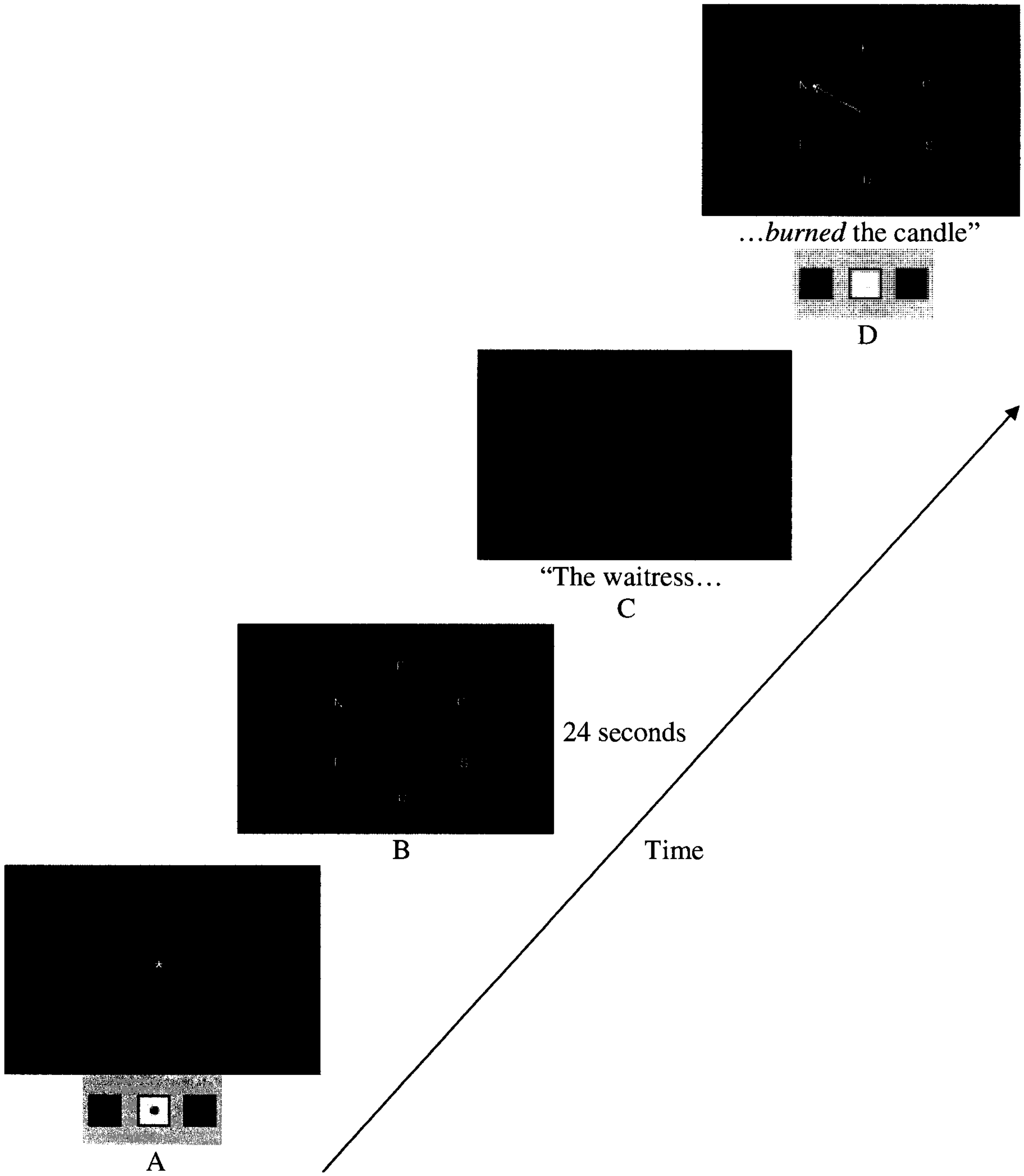


Figure 15. Time sequence of events in Experiment 1c: A) Asterisk fixated until participants pressed the yellow button to begin; B) 6 pictures of objects appeared and center dot was green indicating that participants could freely scan the pictures for 24 seconds; C) pictures disappeared, center dot turned red (which they were asked to fixate), then sentence began; and D) at one of 3 probe points (verb onset, middle of verb or verb offset) 6 pictures of objects re-appeared in the same order, center dot turned green, and if picture of object mentioned in sentence was present, participants were asked to quickly fixate it and press the green button, if not they were asked to press the red button.

Mean Saccade Onset Times to the Target Objects. Mean SOT was defined as the average time taken to initiate a saccade to the target objects, regardless of saccade (i.e., 1st, 2nd, 3rd saccade, etc.). This measure was examined to determine whether verb-specific information could direct the attention system in search of referents of their grammatical objects during sentence comprehension. The prediction was that participants would be quicker at initiating saccades to the target objects as the sentences unfolded, particularly when the sentences contained semantically restrictive verbs, and when given more time to pre-scan the visual displays.

Percent Correct First Saccades to the Target Objects. As the name implies, percent correct first saccades was defined as the percentage of first saccades that were initiated to the target objects. This measure was studied to determine how accurately verb-specific information guided visual attention towards the target objects. This measure was also used to verify whether the obtained mean SOTs were the result of a speed-accuracy trade-off, and to assess participants' memory trace for the arrays of objects. The hypothesis was that as the sentences unfolded participants would make fewer first saccade errors to the target objects, especially in the causative verb condition, and when given more time to pre-scan the visual displays.

Mean Number of Saccades to Reach the Target Objects. Number of saccades was defined as the average number of saccades required to reach the target object. Similar to percent correct first saccades, this measure was used to determine how efficiently verb-specific information directed eye movements towards the target object, and to evaluate the accurateness of participants recall for the arrays of objects. The expectation was that as the sentences unfolded participants would require fewer saccades to reach the referents of the verb's grammatical objects, especially in the causative verb conditions and when given additional time to pre-scan the visual displays.

First Post-Verb Saccades to the Target Objects. First saccade latencies to the target objects were calculated and then compared against their respective onset points within the sentences. This measure was examined to determine whether first saccades were affected by verb type and if they were initiated prior to the utterance of the object nouns in the sentences. The expectation was that if verb-specific information can constrain visual attention at the very early stages of processing, then participants would likely initiate their eye movements to the target objects prior to their noun onset in the accompanying sentences (i.e., anticipatory eye movements).

Cumulative Saccades to the Target Objects. The cumulative number of saccades that were initiated towards the target objects during each 50 ms interval following the onset of the verb was calculated. For each critical point in the sentence (i.e., verb offset, noun onset and noun offset), means for each verb were obtained by dividing the cumulative number of saccades by the total number of trials. This measure was studied to determine whether verb-guided eye movements would be closely time-locked to the utterance of the verb, in particular before the onset of the noun in the sentences. A significant interaction effect between sentence point and verb type was expected, such that the mean number of cumulative saccades to the target objects would rise more quickly at each sentence point for the causative condition than the perception condition, particularly when given more time to pre-scan the visual displays. This prediction was based on the assumption that the interpretation of the verb and the noun phrase occurs incrementally and, therefore, gradually constrain the domain of reference in the scene, especially in the more semantically restrictive causative condition.

Percent First Saccades to the Competitor Objects. The percentages of first saccades to the different competitor objects (i.e., phonological, semantic, color, shape and unrelated) were calculated and averaged across verb and probe point conditions. This measure was examined to determine whether certain types of competitor objects, which share a feature with the target object, attract attention at different points within the processing of the sentence. The prediction was that competitor objects that shared linguistic and semantic properties with the target objects would attract attention in the earlier stages of sentence processing, particularly when given less time to pre-scan the visual displays.

1.5.1 Mean Saccade Onset Times to Target Objects

Table 1 summarizes the results of the analyses conducted on the mean SOT values across the three experiments, and Figure 16 presents the mean SOTs per verb type across probe points for each experiment. The top panel of Table 1 shows, as expected, a statistically significant difference between probe points for each experiment was obtained. Bonferroni post-hoc pairwise comparisons indicated that for Experiment 1a the difference between probe points lay between the noun onset and noun offset probe points, and noun onset and noun offset+200 ms probe point conditions; between verb onset and verb offset probe point conditions for Experiment 1b; and between each verb probe point for Experiment 1c. Therefore, as the delay between the onset of the sentence and the green dot increased, the time to initiate a saccade to the target object decreased (see Figure 16). Faster SOT at later probe points suggests that by the onset of the green dot

Table 1. Summary of analyses conducted on mean saccade onset times (SOT) and mean SOT corrected for verb length, across all three experiments (Noun – 6 sec; Verb – 6 sec; and Verb – 24 sec).

Study	Measure	Factor	Analysis	p	η^2
Noun	Mean SOT	Verb Type	$F1(1,28)= 0.153$	.699	.005
(6 sec)			$F2(1,22)= 0.015$	.904	.001
		Probe Point	$F1(2,56)= 7.25$	.002 *	.206
			$F2(2,44)= 3.87$	.028 *	.149
Verb	Mean SOT	Verb Type	$F1(1,26)= 13.56$	.001 *	.343
(6 sec)			$F2(1,22)= 8.37$	.008 *	.276
		Probe Point	$F1(2,52)= 4.76$	.013 *	.155
			$F2(2,44)= 4.73$	.014 *	.177
Verb	Mean SOT	Verb Type	$F1(1,17)= 22.14$	< .001 *	.566
(24 sec)			$F2(1,23)= 23.62$	< .001 *	.507
		Probe Point	$F1(2,34)= 23.76$	< .001 *	.583
			$F2(2,46)= 47.29$	< .001 *	.673
Verb	Mean SOT –	Verb Onset	$F2(1,21)= 4.07$	.057	.162
(6 sec)	Accounting for	Verb Middle	$F2(1,21)= 1.82$	.192	.080
	Verb Length	Verb Offset	$F2(1,21)= 0.197$	.661	.009
Verb	Mean SOT –	Verb Onset	$F2(1,22)= 3.48$	.075	.137
(24 sec)	Accounting for	Verb Middle	$F2(1,22)= 11.36$	.003 *	.341
	Verb Length	Verb Offset	$F2(1,22)= 6.62$	.017 *	.231

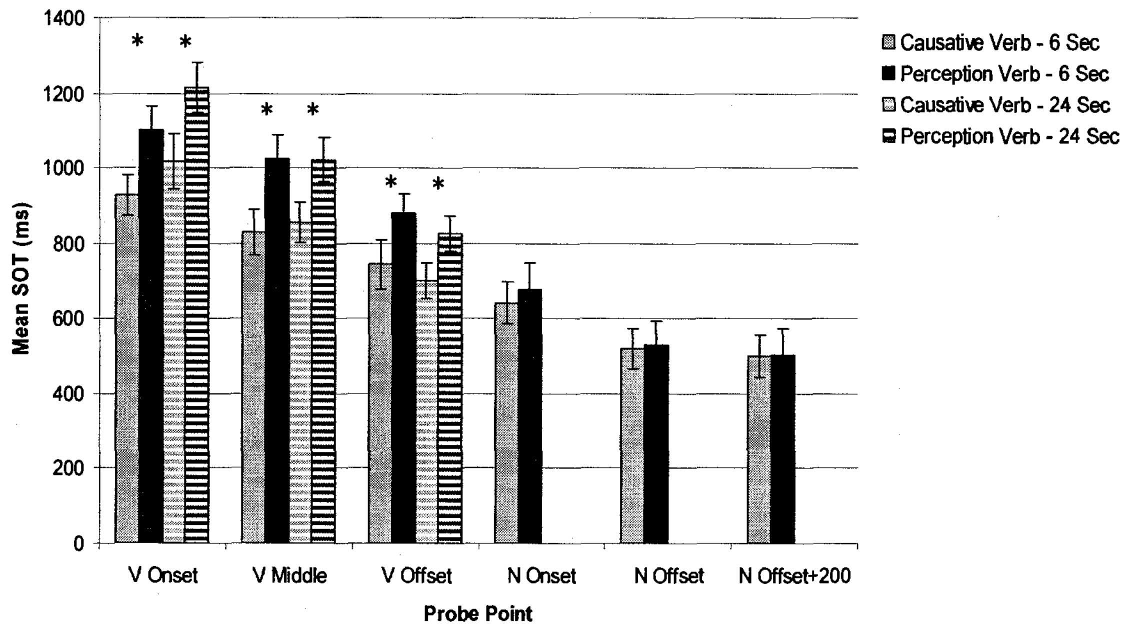


Figure 16. Mean saccade onset times (SOT) to target objects, per verb type across probe points in Experiment 1a, 1b and 1c. Error bars indicate standard error of the mean.

participants had more time to complete the linguistic analysis and, thus, began their search at a point when greater certainty about the identity of the target noun was achieved. In addition, as there was no statistically significant difference between the noun offset and noun offset+200 ms probe point conditions, it appears that the processing and programming of the target object was completed some time before the offset of the word. Interestingly, when comparing mean SOT values between Experiment 1b and 1c the increased scanning time provided in Experiment 1c did not significantly decrease average saccade onset time to the target object, $F1(1,43)= 0.205$, $p= .653$, $\eta^2= .005$; $F2(1,45)= 0.025$, $p= .876$, $\eta^2= .001$. These results, therefore, indicate that the lengthy mean SOT values found in Experiment 1b were not simply the result of a poor internal representation of the visual display.

Finally, as indicated in Table 1, there was a statistically significant difference between verb types when probing at the verb (Experiment 1b and 1c), with shorter mean SOTs in the causative verb condition. This effect, however, was no longer present when probing during the presentation of the noun. In fact, Figure 16 demonstrates that the causative verb class advantage decreased as the sentences unfolded. These results, therefore, seem to indicate that the mediating effects of semantically restrictive verbs in guiding eye-movements only occurs at the earlier stages of sentence processing, well before the onset of the post-verbal noun.

1.5.2 Mean Saccade Onset Times to Target Objects Controlling for Verb Length

Another variable, however, may be responsible for the reported difference in mean SOTs between these two verb classes. Table 2 summarizes the mean duration of the verbs across the two conditions, the mean delay between verb offset and noun onset,

Table 2. Word durations for the causative and perception sentences (ms)

Duration	Causative	Perception	Difference in ms (Perception – Causative)
Verb	455	530	75
Post-verbal break + Determiner	233	236	3
Target Noun	435	421	14
Verb + Break + Determiner	688	766	78

the mean duration of the target noun, and the mean delay between onset of the verb and onset of the target noun. A paired samples t-test revealed that mean verb duration $t(23)= -2.69$, $p= .013$ and mean delay between onset of the verb and onset of the target noun $t(23)= -2.62$, $p= .015$ differed significantly across the two verb classes. By contrast, mean delay between verb offset and noun onset $t(23)= -0.219$, $p= .829$; and mean target noun duration $t(23)= 1.78$, $p= .089$ did not differ significantly across verb classes. Verb length, therefore, may have acted as a confounding variable in Experiments 1b and 1c. Because causative verbs on average required less time to utter than the perception verbs, the target nouns tended to be heard sooner within the sentence, and thus may have been responsible for the observed quicker SOTs.

Consequently, mean SOTs were re-analyzed using three one-way repeated measures ANCOVAs, using the difference between each point within the verb (i.e., onset, middle and offset) and the onset of the noun as a covariate. As demonstrated in the bottom panel of Table 1, controlling for verb length resulted in a significant reduction in F values, suggesting that some of the advantage found for the causative verbs could be accounted for by the significant delay differences between the two verb classes. In Experiment 1c, however, even when accounting for this difference, participants were faster at initiating a saccade to the target object when it was preceded by a causative verb. Furthermore, as shown in Table 2, the average delays between verb probe points and noun onset for the causative verb sentences were 688, 461, and 233ms, respectively. Assuming that a saccade requires approximately 200 ms to program (Matin et al., 1993), these saccades were initiated *after* noun onset (38, 169, and 311ms respectively) yet before noun offset. Thus, on average, the more semantically restrictive causative verbs did not result in anticipatory eye-movements, but rather appear to have accelerated the recognition point (i.e., point at which a word becomes recognizable) of the target noun.

1.5.3 Percent Correct First Saccades to Target Objects

Table 3 summarizes the results of the analyses conducted on the percent correct first saccades across the three experiments, and Figure 17 presents the percent correct first saccades per verb type across probe points for each experiment. As demonstrated in the top panel of Table 3, results were not as predicted. There was no statistically significant effect of probe point, and with the exception of the by subject's analysis for Experiment 1b, no statistically significant effect of verb type was observed. These results suggest that the decrease in mean SOT values across probe points and in the causative verb condition (during the verb only) were not due to a speed-accuracy tradeoff (i.e., that they were slower at the earlier probe points and for the perception verbs because they made more first saccade errors). Rather, two possible explanations for these SOT values are that: (1) participants produced more saccades before reaching the target object in the earlier probe point and perception verb condition or (2) the speed of the saccades and/or the length of fixations became faster as participants were probed later within the sentence, as well as in the causative verb condition. The latter appears to be the most likely explanation, as discussed in the next section.

As shown in Figure 17, a large number of errors were obtained for the first saccades in Experiment 1a and 1b (65% and 70%). Although performance was somewhat poor, the results demonstrated that participants still had some memory trace for the location of the objects in the display as they performed better than chance (i.e., 16.67%). When participants were given extra time to scan the visual display they still

Table 3. Summary of analyses conducted on percent correct first saccades and number of saccades to reach the target object, across all three experiments (Noun – 6 sec; Verb – 6 sec; and Verb – 24 sec).

Study	Measure	Factor	Analysis	p	η^2
Noun	Correct 1st Saccade	Verb	$F1(1,29)= 0.173$	.681	.006
(6 sec)			$F2(1,23)= 0.081$	.778	.004
		Probe Point	$F1(2,58)= 0.413$	.664	.014
			$F2(2,46)= 0.807$	.453	.034
Verb	Correct 1st Saccade	Verb	$F1(1,28)= 4.73$	.038 *	.144
(6 sec)			$F2(1,23)= 3.22$	.086	.123
		Probe Point	$F1(2,56)= 0.285$	.753	.010
			$F2(2,46)= 0.075$	.927	.003
Verb	Correct 1st Saccade	Verb	$F1(1,17)= 1.26$	.277	.069
(24 sec)			$F2(1,23)= 1.54$	.227	.063
		Probe Point	$F1(2,34)= 0.743$	.483	.042
			$F2(2,46)= 0.741$	.482	.031
Noun	Number of Saccades	Verb	$F1(1,28)= 0.434$	.515	.015
(6 sec)			$F2(1,23)= 1.11$	.303	.046
		Probe Point	$F1(2,56)= 1.58$	.216	.053
			$F2(2,46)= 1.03$	.365	.043
Verb	Number of Saccades	Verb	$F1(1,28)= 1.97$	.172	.066
(6 sec)			$F2(1,23)= 4.52$	.044 *	.164
		Probe Point	$F1(2,56)= 0.092$	.912	.003
			$F2(2,46)= 0.225$	.799	.010

Verb	Number of Saccades	Verb	$F1(1,17)= 3.58$	.076	.174
(24 sec)			$F2(1,23)= 2.52$	.126	.099
		Probe Point	$F1(2,34)= 0.283$	.667	.016
			$F2(2,46)= 0.363$	.698	.016

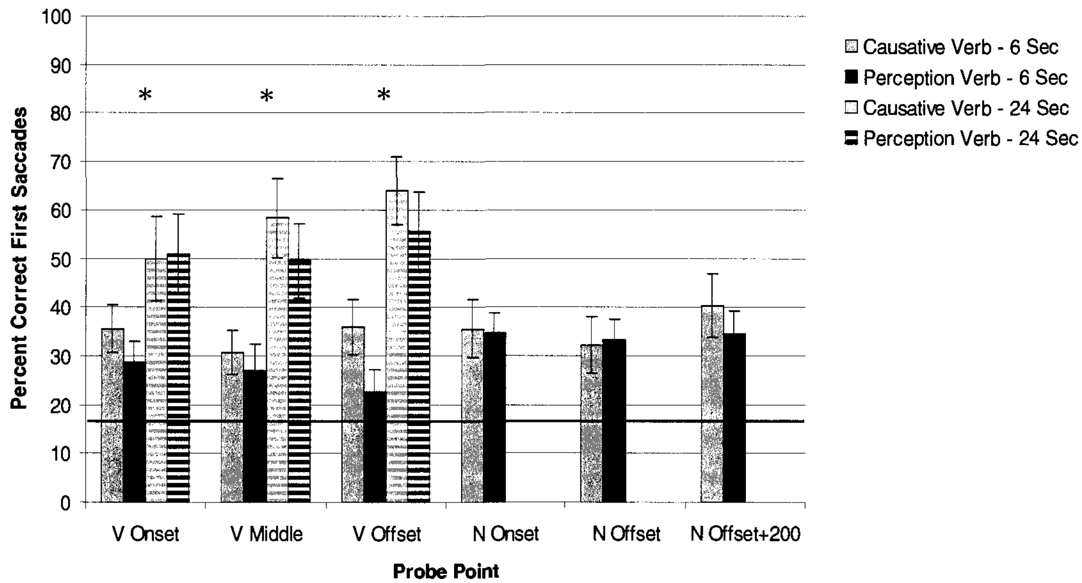


Figure 17. Percent correct first saccades to target objects, per verb type across probe points in Experiment 1a, 1b and 1c. The red line depicts chance level.

made quite a few first saccade errors (approximately 46%), however, less than in Experiment 1b. In fact, when comparing performance across these two studies participants made significantly fewer first saccade errors $\underline{F1}(1,45)= 20.89$, $\underline{p}< .001$, $\eta_p^2 = .317$, $\underline{F2}(1,46)= 37.38$, $\underline{p}< .001$, $\eta_p^2 = .448$. Therefore, increased scanning time appears to have produced a more stable and accurate internal representation of the visual display, which resulted in participants making fewer first saccade errors. This more accurate internal representation, however, does not appear to have led to quicker saccades to the target object, as no significant difference in mean SOTs was found between Experiment 1b and 1c in the previous section. Rather, participants who had the opportunity to consolidate into memory the identity and location of the objects in the display appear to have delayed their first saccades, thus resulting in fewer first saccade errors (i.e., a speed-accuracy tradeoff).

1.5.4 Number of Saccades to Reach Target Objects

Table 3 summarizes the results of the analyses conducted on the number of saccades across the three experiments, and Figure 18 presents the number of saccades to reach the target object per verb type across probe points for each experiment. As demonstrated in Table 3, a similar pattern of results emerged across experiments, with no statistically significant effect of probe point or verb class, with the exception of the items analysis in Experiment 1b. These results, therefore, appear to support the conclusion mentioned in the previous sections; that the decrease in mean SOTs across probe points and in the causative verb condition (when probed at the verb) were not the result of a speed-accuracy tradeoff, but rather quicker saccades and/or shorter fixations in these conditions. Furthermore, as shown in Figure 18, when given 6 seconds to initially scan

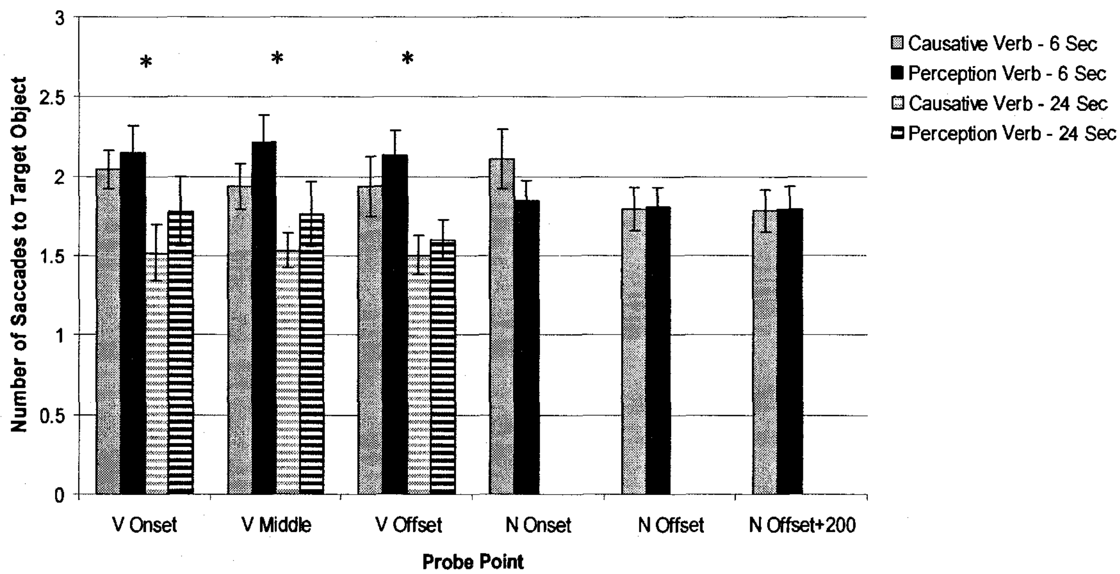


Figure 18. Average number of saccades to reach target objects, per verb type across probe points in Experiment 1a, 1b and 1c.

the display, participants on average required fewer than two saccades to reach the target object which indicates that participants had some memory trace of the location of the objects. When given more time to examine the visual display in Experiment 1c, however, participants required significantly fewer saccades to reach the target object, compared to Experiment 1b, $\underline{F1}(1,45) = 10.72$, $\underline{p} = .002$, $\underline{\eta}_p^2 = .192$, $\underline{F2}(1,46) = 17.32$, $\underline{p} < .001$, $\underline{\eta}_p^2 = .274$. Thus it appears that the additional scanning time proved helpful in producing a more stable and accurate internal representation of the visual display. This in turn reduced first saccade errors and number of saccades required to reach the target object.

As discussed in the previous section, however, this more accurate internal representation did not significantly decrease mean saccade onset times, suggesting that participants may have behaved differently across studies. The greater number of saccades in Experiment 1b lends further support to the idea that a less stable and accurate internal representation led participants to search for the target object sooner, and possibly more easily distracted by the competitor objects and prone to interference from subsequent visual information. On the other hand, with a more stable and accurate internal representation of the visual display (Experiment 1c), participants' saccades were delayed until the target object became uniquely identifiable, and as a result made fewer first saccade errors.

1.5.5 First Post-Verb Saccades to the Target Objects

The mean SOTs reported above reflected participants' performance across multiple saccades. These values, however, are confounded by the distance between the positioning of the first incorrect saccade relative to the target object (e.g., next to the target object vs. across). Although the positioning of the target object was counterbalanced across trials and participants, due to the design of the study not all possible combinations were possible (i.e., 6! = 720). In order to eliminate these possible effects only SOTs for first correct saccades to the target objects were examined. As was demonstrated in previous sections, however, participants made numerous first saccade errors (approximately 65%; 70%; and 46% for Experiment 1a, 1b, and 1c, respectively), leading to many missing data (approximately 23%; 30%; 16%, respectively). In order to conduct an analysis with adequate power the missing values were replaced using the Expectation Maximization (EM) approach (see Schafer, 1997). Although this method is commonly used to replace missing values, the results should still be interpreted with caution.

Table 4 summarizes the results of the analyses conducted on the correct first SOTs across the three experiments, and Figure 19 presents the first SOTs per verb type and across probe points for each experiment. As shown in the top panel of Table 4, there was a statistically significant difference between probe points for each experiment. Bonferroni post-hoc pairwise comparisons indicate that the difference between probe points lay between each probe point for Experiment 1a, 1b and 1c. As shown in Figure 19, the same pattern of results emerged as for the mean SOT data; participants were quicker to initiate a saccade to the target object as the delay between the onset of the noun and the green dot increased. Moreover, when comparing first SOT values between Experiment 1b and 1c, the increased scanning time provided in Experiment 1c significantly increased correct first saccade onset times to the target objects, $F1(1,34)= 11.36$, $p= .004$, $\eta_p^2 = .401$; $F2(1,46)= 18.59$, $p< .001$, $\eta_p^2 = .288$. The slower first SOT values appear to lend further credence to the theory mentioned above, that with a more

Table 4. Summary of analyses conducted on correct first saccade onset times (SOT) and first SOT correcting for verb length, across all three experiments (Noun – 6 sec; Verb – 6 sec; and Verb – 24 sec).

Study	Measure	Factor	Analysis	p	η^2
Noun	First SOT	Verb	$F1(1,29) = 0.054$	.817	.002
(6 sec)			$F2(1,23) = 0.225$	.640	.010
		Probe Point	$F1(2,58) = 22.03$	< .001 *	.432
			$F2(2,46) = 16.53$	< .001 *	.418
Verb	First SOT	Verb	$F1(1,28) = 5.00$	.033 *	.152
(6 sec)			$F2(1,23) = 2.50$	.128	.098
		Probe Point	$F1(2,56) = 16.49$	< .001 *	.371
			$F2(2,46) = 10.13$	.001 *	.306
Verb	First SOT	Verb	$F1(1,17) = 11.36$	.004 *	.401
(24 sec)			$F2(1,23) = 13.68$	.001 *	.373
		Probe Point	$F1(2,34) = 27.30$	< .001 *	.616
			$F2(2,46) = 23.53$	< .001 *	.506
Verb	First SOT –	Verb Onset	$F2(1,22) = 0.605$	.445	.027
(6 sec)	Accounting for	Verb Middle	$F2(1,22) = 3.74$	.066	.145
	Verb Length	Verb Offset	$F2(1,22) = 0.380$	.544	.017
Verb	First SOT –	Verb Onset	$F2(1,22) = 1.24$	.279	.053
(24 sec)	Accounting for	Verb Middle	$F2(1,22) = 1.54$	.228	.065
	Verb Length	Verb Offset	$F2(1,22) = 6.43$	.019 *	.226

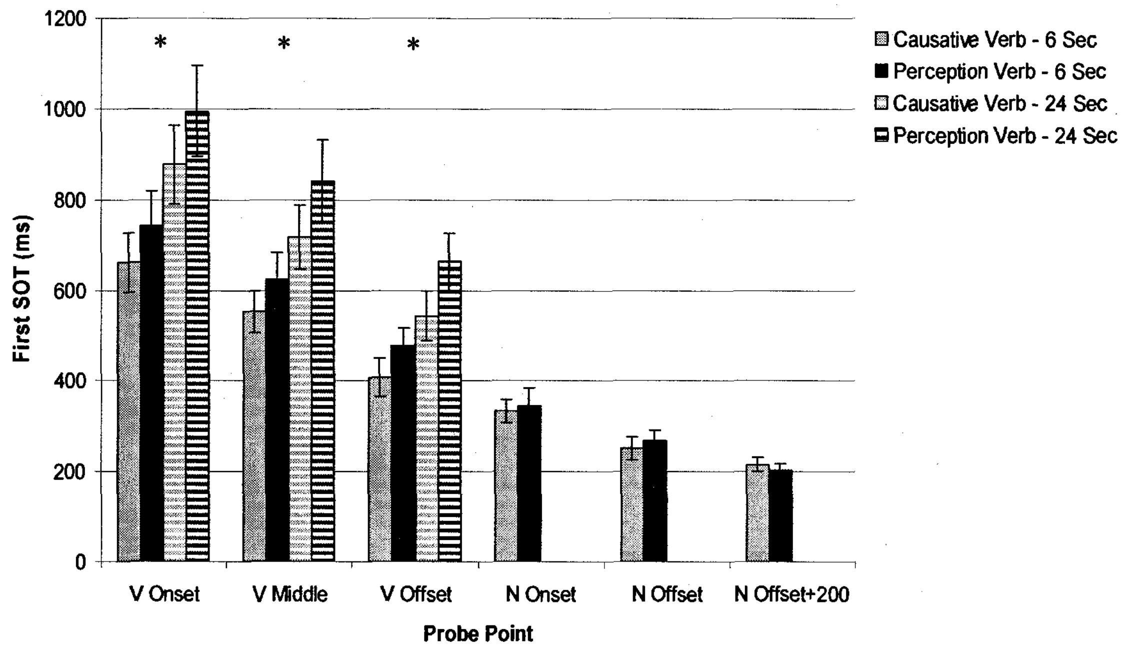


Figure 19. First saccade onset times to target objects, per verb type across probe points in Experiment 1a, 1b and 1c.

stable and accurate internal representation participants tended to delay their saccades until the target noun within the sentence became uniquely identifiable.

Examining Table 4 once again, there was a statistically significant difference between verb types when probing at the verb (Experiment 1b and 1c), with shorter correct first SOTs in the causative verb condition. This effect, however, was no longer present when probing during the presentation of the noun. In fact Figure 19 demonstrates that the difference between verb classes decreased as the sentence unfolded. These results, therefore, suggest that the mediating effects of the semantically restrictive verbs on guiding attention to the visual referent of the verb's direct object occurs at the earlier stages of sentence processing. As mentioned previously, however, differences in verb length between these two verb classes may have acted as a confounding variable. Consequently, correct first SOTs were re-analyzed using three one-way repeated measures ANCOVAs, using the difference between each point within the verb (i.e., onset, middle and offset) and the onset of the noun as a covariate.

1.5.6 First Post-Verb Saccades to Target Object Controlling for Verb Length

As demonstrated in the bottom panel of Table 4, controlling for verb length resulted in a significant reduction in F values, suggesting that some part of the advantage found in the causative verb condition could be accounted for by the significant delay differences between the two verb classes. In Experiment 1c, however, even when accounting for this difference, participants were faster at initiating a saccade to the target object when it was preceded by a semantically restrictive causative verb.

1.5.7 Anticipatory Eye Movements

To determine whether these eye movements were anticipatory in nature (i.e., programmed before the onset of the noun), in Figure 20 correct first saccades were plotted relative to noun onset for each verb type across probe points for Experiment 1b and 1c. As can be seen in Figure 20 all values are positive, indicating that on average saccades were initiated after the onset of the noun. Saccades were initiated 32, 78, and 321 ms after the onset of the noun in the causative condition, compared to 156, 211, and 269 ms in the perception condition, for the verb onset, middle and offset conditions respectively.

Assuming that a saccade requires approximately 200 ms to program (Matin et al., 1993), however, these results suggest that early within the sentence participants with a less stable and accurate internal representation appear to have used the semantically restrictive nature of the causative verbs to make anticipatory eye movements to the referent of the complement of the verb. Unexpectedly, participants probed early within the *perception* verb sentences also made anticipatory eye-movements to the target objects. As the perception verbs provided little help with regards to restricting the set of possible noun referents within the visual display, it appears that it was the semantic restrictions afforded by the agents within the majority of the sentences (i.e., 23/24) that allowed for these anticipatory eye-movements to occur (e.g., "The *mechanic* scrutinized the *wheel*").

In contrast, when probed later within the sentence and/or when a more stable and accurate internal representation had been acquired, participants saccades were delayed until after the onset of the noun. Saccades were initiated 383, 403, and 453 ms after noun onset in the causative condition, compared to 474, 506, and 590 ms in the perception condition, for the verb onset, middle and offset conditions respectively. In Experiment

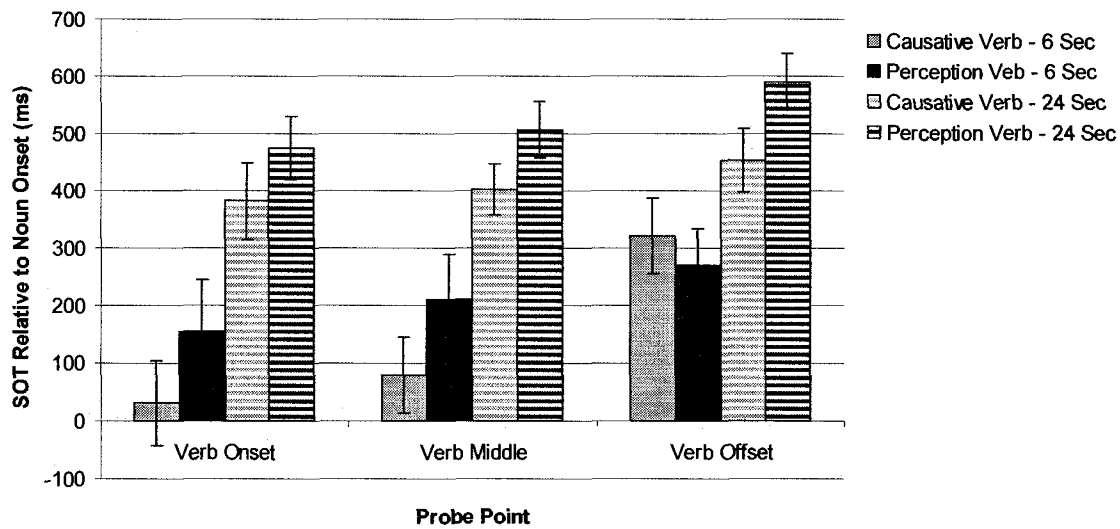


Figure 20. First saccade onset times to target objects relative to noun onset, per verb type across probe points in Experiment 1b and 1c.

1c, the quicker SOT values for the causative verb condition suggest that the semantically restrictive nature of this verb class may have also resulted in recognition of the target noun before its uniqueness point (i.e., point where the initial sequence is common only to that word).

Another explanation for the observed pattern of results is that the task demands (yes/no judgment task and "move your eyes as quickly as possible") may have "unnaturally" pushed participants to predict the identity of the target noun – with only moderate success as evidenced by the relatively high error rates. If this assumption is true, then saccades towards incorrect objects would be expected to have shorter SOTs than saccades to correct target objects, because the chance of correctly identifying the target noun, and thus avoiding a saccade towards an incorrect target object, is greater the more of the unfolding sentence has been heard. The data in Figure 21, which demonstrates average SOTs for incorrect first saccades relative to noun onset, appear to support this hypothesis as nearly all saccades were planned within the presentation of the verb, many early within the verb, and launched sooner than correct first saccades. Even in the perception verb condition, in which the verb provided little help in constraining the domain of reference, participants anticipated the target noun by means of the semantic restrictions afforded by the agent. An additional study, in which the probe occurs during the presentation of the agent, however, would be required to substantiate this hypothesis.

Furthermore, similar to correct first saccades, as the delay between the green dot and noun onset decreased, SOT for incorrect first saccades increased (especially when given 24 seconds to initially scan the display). It remains unclear at this point whether the delay in launching a saccade is strategic (i.e., wait a little to avoid wasting time in

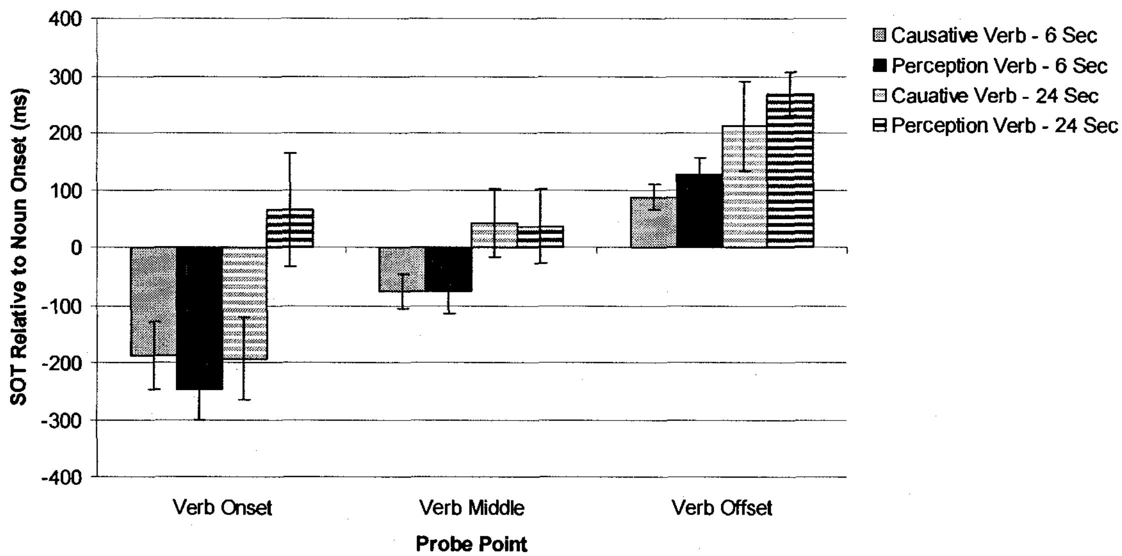


Figure 21. Incorrect first saccade onset times to target objects relative to noun onset, per verb type, across probe points in Experiment 1b and 1c.

programming and initiating a saccade to an incorrect object, to then start over the process until the correct object is reached), and/or as a function of the strength of consolidation into memory (i.e., fleeting and therefore necessitating quick responding, or stable and thus allowing for more cautious responding).

1.5.8 Cumulative Saccades to the Target Objects

Finally, based on data analysis techniques suggested by Altmann and Kamide (2004), the mean number of cumulative saccades to the target objects for each verb class across probe points was plotted in Figure 22. For each verb type and probe point, the average number of cumulative saccades to the target objects was calculated for each 50-ms interval from the onset of the green dot until 2500 ms. The means do not sum to 1.0 by trial offset because in 10.75%, 8.76% and 5.09% of trials, in Experiment 1a, 1b and 1c respectively, participants did not initiate a saccade towards the target object.
Sentence position, was added to the analysis as an additional factor to produce two three-way repeated measure ANOVAs. Differences in cumulative saccade means between the two verb classes were analyzed at two critical sentence positions for Experiment 1a (noun offset and noun offset+200 ms) and three critical sentence positions for Experiment 1b and 1c (noun offset, verb onset, and verb offset). As the speaker's rate of speech and the length of the individual words in each sentence were different, the critical sentence positions for each sentence also differed. For each participant the cumulative saccade means were, therefore, taken from the different 50-ms bins corresponding to these critical positions. Although they occurred at different points in time, they still corresponded to the same linguistic markers (i.e., the end points of the verb and noun, and the onset of the noun).

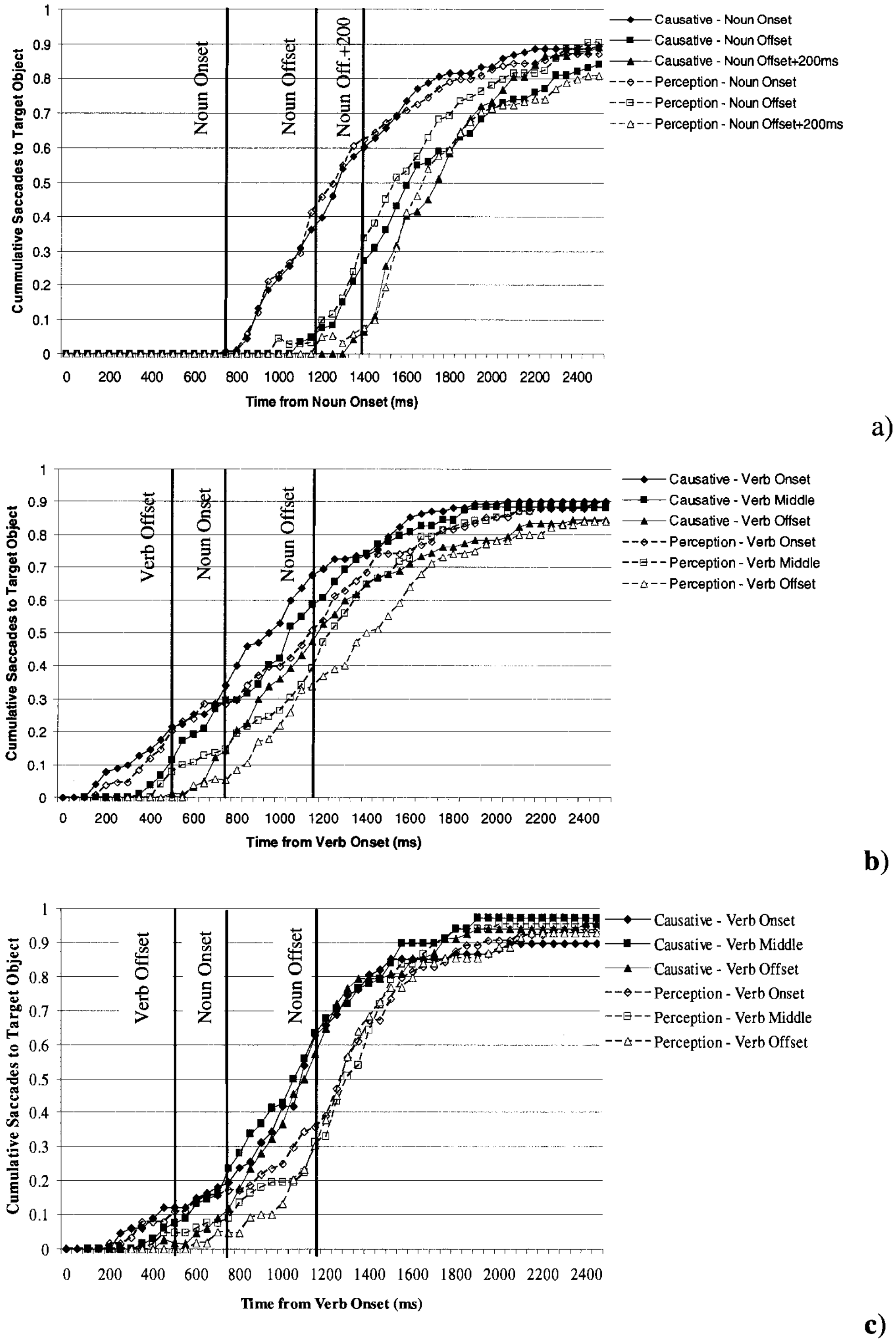


Figure 22. Summary of cumulative saccades to target object across all 3 experiments: a) probed at the noun (6 sec); b) probed at the verb (6 sec); and c) probed at the verb (24 sec). Each tick mark on the X-axis refers to one 50ms bin, and each dotted line to a single condition. The vertical lines mark the average onset and offsets of the noun and verb.

Table 5. Summary of analyses conducted on cumulative saccade means at critical sentence positions, across all three experiments (Noun – 6 sec; Verb – 6 sec; and Verb – 24 sec).

Study	Measure	Factor	Analysis	p	η^2
Noun	Cumulative	Position	$F1(1,29)= 66.41$	< .001 *	.696
(6 sec)	Saccades		$F2(1,23)= 125.67$	< .001 *	.845
		Verb	$F1(1,29)= 1.49$	.232	.049
			$F2(1,23)= 0.167$	.686	.007
		Probe Point	$F1(1,29)= 74.30$	< .001 *	.719
			$F2(1,23)= 62.16$	< .001 *	.730
Verb	Cumulative	Position	$F1(2,56)= 172.36$	< .001 *	.860
(6 sec)	Saccades		$F2(2,46)= 150.26$	< .001 *	.867
		Verb	$F1(1,28)= 5.27$	.029 *	.158
			$F2(1,23)= 3.14$	.090	.120
		Probe Point	$F1(2,56)= 14.71$	< .001 *	.344
			$F2(2,46)= 11.88$	< .001 *	.341
Verb	Cumulative	Position	$F1(2,56)= 94.15$	< .001 *	.847
(24 sec)	Saccades		$F2(2,46)= 107.75$	< .001 *	.824
		Verb	$F1(1,29)= 5.50$	.031 *	.244
			$F2(1,23)= 17.19$	< .001 *	.428
		Probe Point	$F1(2,56)= 2.92$	.085	.147
			$F2(2,46)= 4.09$	.023 *	.151

As indicated in Table 5, there was a statistically significant difference between probe points in Experiment 1a and 1b. Bonferroni post-hoc pairwise comparisons indicate that the difference between probe points lay between the noun onset and noun offset probe points for Experiment 1a, and between each verb probe point condition for Experiment 1b. By contrast, in Experiment 1c, there was only a marginally significant effect of probe point. In addition, there was a statistically significant difference between sentence positions for each experiment. Bonferroni post-hoc pairwise comparisons indicate that the differences lay between each sentence position in Experiment 1a, 1b and 1c. As these results and the slopes in Figure 22 demonstrate, participants made increasingly more saccades to the target object as the sentence unfolded. The steeper slope of the noun offset+200 ms curve in Experiment 1a appears to indicate that participants had already identified and programmed the saccade to the target object before the onset of the green dot. When initially given more time to scan the visual display (Experiment 1c) the three slopes for the causative verb condition converged and became nearly identical midway within the presentation of the noun, and by noun offset for the perception verb condition. These convergence points possibly signal when the noun became recognizable (minus approximately 200 ms to program the saccade) within the sentences of each verb condition.

Furthermore, as indicated in Table 5 there was a statistically significant difference between verb classes when probing at the verb (Experiment 1b and 1c); with a greater number of cumulative saccades to the target object in the causative verb condition. This effect, however, was no longer present when probing at the noun (Experiment 1a). There was also a significant interaction between verb type and sentence position for Experiment

1b ($\underline{F1}(2,56)= 5.41$, $\underline{p}= .018$, $\eta_p^2 = .162$; $\underline{F2}(2,46)= 4.91$, $\underline{p}= .012$, $\eta_p^2 = .176$) and Experiment 1c ($\underline{F1}(2,56)= 4.19$, $\underline{p}= .047$, $\eta_p^2= .198$; $\underline{F2}(2,46)= 17.44$, $\underline{p}< .001$, $\eta_p^2 = .431$). As the noun phrase was processed incrementally and restrictive semantic information was provided by the verb, the mean number of cumulative saccades to the target object increased more quickly for the causative condition than the perception condition. In most cases, however, the effect sizes were quite small. In addition, since verb length differences in each sentence condition was taken into account when calculating the cumulative saccade means at each critical sentence position (see second paragraph of this section for further detail), the results lend support to the conclusions made following the aforementioned ANCOVA analyses – that significant differences in verb length between the two classes of verbs accounted for some of the obtained verb effects.

A closer look at Figure 22b and 22c also shows that the average number of cumulative saccades to the target object in the causative verb condition significantly increased around the onset of the noun and reached a mean of approximately 0.70 by noun offset. Assuming that a saccade requires approximately 200 ms to program (Matin et al., 1993), the slopes would suggest that in the causative verb class condition some anticipatory eye movements were planned during the presentation of the verb. Saccades were launched before the onset of the noun in only 27.0%, 21.6% and 9.77% of the causative trials, compared to 28.7%, 11.2% and 5.17% of the perception trials, for the verb onset, middle, and offset conditions respectively. In contrast, for the perception verb condition this drastic increase in cumulative saccade means occurred more towards the offset of the noun. Because the semantically less restrictive perception verbs provided little help in constraining the domain of reference, participants had to wait until the utterance of the noun before initiating the appropriate saccade. Yet some anticipatory eye movements did occur before the onset of the noun, and as discussed above, may be the result of the semantic restrictions afforded by the agent that provided some additional guidance. Finally, the delay in initiating a saccade for the participants with a more stable and accurate internal representation of the visual scene can be observed in Figure 22c by the less steep slopes before noun onset, as compared to Figure 22b. After this point the average number of cumulative saccades in Experiment 1c began to rise much more quickly than in Experiment 1b. This apparent delay in launching saccades to the target object is further evidenced by the small number of anticipatory eye movements, which only occurred in 20.4%, 17.1% and 5.56% of the causative trials, compared to 17.1%, 11.1% and 2.78% of the perception trials, for the verb onset, middle and offset conditions, respectively.

1.5.9 Competitor Objects

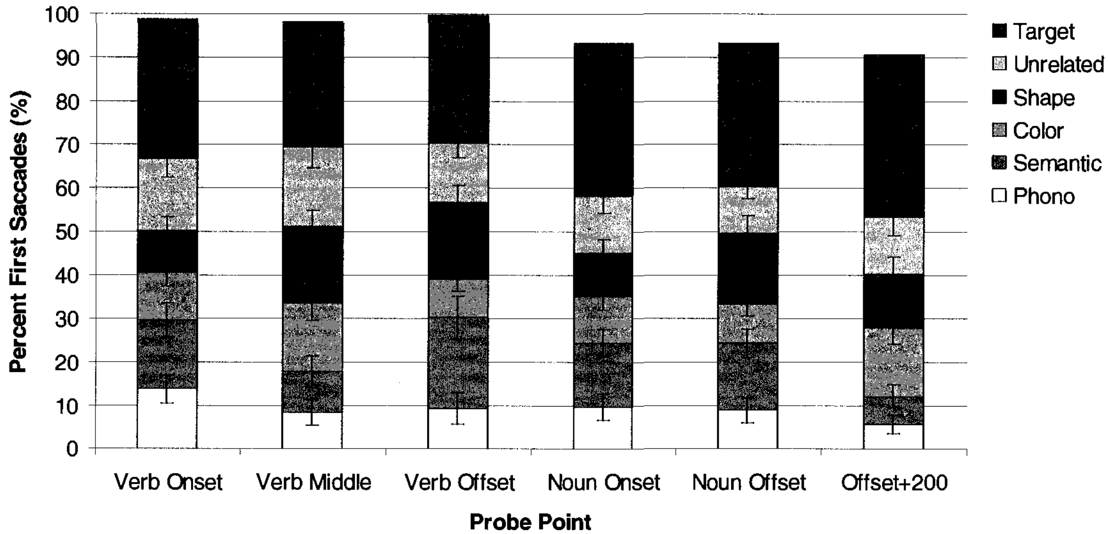
An additional factor, competitor type (i.e., phonological, function, color, shape, and unrelated), was added to the analysis to produce two three-way repeated measure ANOVAs. Table 6 summarizes the results of the analyses conducted on the first saccades to competitor objects across the three experiments, and Figure 23a presents the percentage of first saccades to each type of competitor object across probe points for Experiment 1a and 1b. Figure 23b presents the percentages when participants were given 24 seconds to pre-scan the visual display (i.e., Experiment 1c). Because there were no statistically significant verb effects or verb interactions, the data in Figure 23a and 23b were collapsed across verb classes. In some conditions the total did not add to 100

Table 6. Summary of analyses conducted on percent first saccades to each competitor type and percent first saccades to each competitor type with incongruent items removed, across all three experiments (Noun – 6 sec; Verb – 6 sec; and Verb – 24 sec).

Study	Measure	Factor	Analysis	p	η^2
Noun	Competitors	Type	$F1(4,116) = 1.85$	.124	.060
(6 sec)			$F2(4,92) = 1.11$	.357	.046
		Verb	$F1(1,29) = 0.265$	.611	.009
			$F2(1,23) = 0.059$	.810	.003
		Probe Point	$F1(2,58) = 0.923$	.403	.031
			$F2(2,46) = 1.09$	.337	.045
Verb	Competitors	Type	$F1(4,112) = 2.69$	.035 *	.088
(6 sec)			$F2(4,92) = 1.14$	.342	.047
		Verb	$F1(1,28) = 4.06$	.054	.127
			$F2(1,23) = 2.64$	.118	.118
		Probe Point	$F1(2,56) = 0.364$	.687	.013
			$F2(2,46) = 0.087$	.917	.004
Verb	Competitors	Type	$F1(4,68) = 5.13$	.001 *	.232
(24 sec)			$F2(4,92) = 3.26$	.034 *	.124
		Verb	$F1(1,17) = 0.721$	.408	.041
			$F2(1,23) = 2.09$	.162	.083
		Probe Point	$F1(2,34) = 0.751$	.480	.042
			$F2(2,46) = 0.630$	.537	.027
Noun	Competitors	Type	$F1(4,116) = 2.63$	.038 *	.083
(6 sec)	Incongruent		$F2(4,32) = 2.70$	.080	.253

	Removed	Verb	F1(1,29)= 0.543	.467	.018
			F2(1,8)= 0.500	.500	.059
		Probe Point	F1(2,58)= 0.763	.471	.026
			F2(2.16)= 0.329	.725	.039
Verb	Competitors	Type	F1(4,112)= 2.57	.042 *	.084
(6 sec)	Incongruent		F2(4,32)= 0.219	.926	.027
	Removed	Verb	F1(1,28)= 2.55	.122	.083
			F2(1,8)= 5.60	.046 *	.412
		Probe Point	F1(2,56)= 0.299	.743	.011
			F2(2,16)= 0.873	.437	.098
Verb	Competitors	Type	F1(4,68)= 7.97	<.001*	.319
(24 sec)	Incongruent		F2(4,32)= 4.06	.009 *	.336
	Removed	Verb	F1(1,17)=0.762	.395	.043
			F2(1,8)=0.341	.576	.041
		Probe Point	F1(2,34)=1.47	.244	.080
			F2(2,16)=0.004	.996	.001

a)



b)

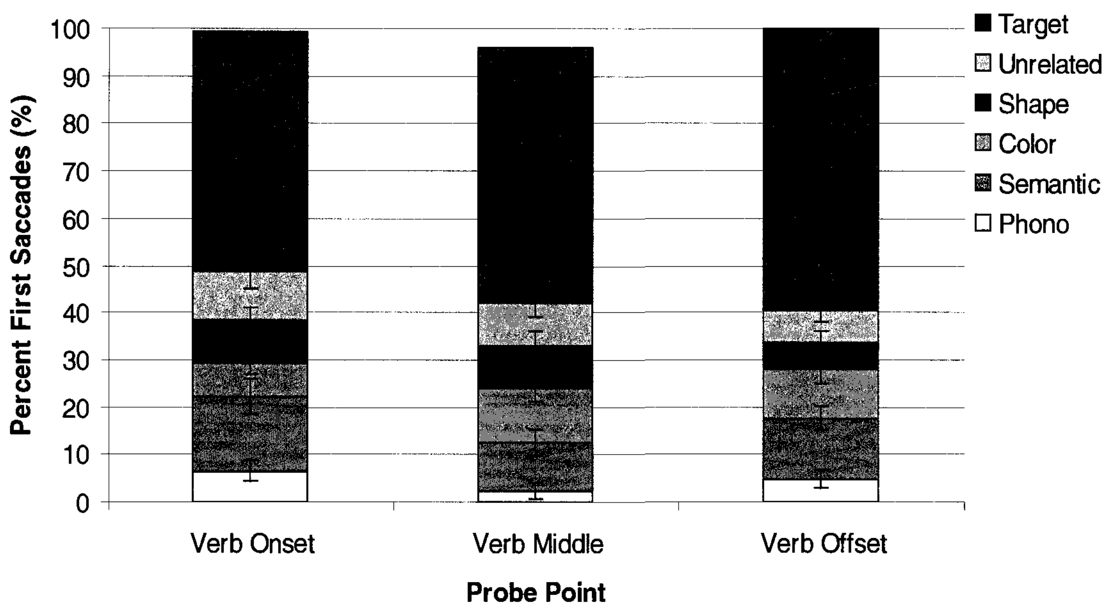


Figure 23. Percentage of first saccades made to each competitor type, across probe points, for a) Experiment 1a and 1b, and b) Experiment 1c

percent because some saccades were launched to empty regions between two pictures of objects and, therefore, could not be coded.

As indicated in the first panel of Table 6, in Experiment 1a there were no statistically significant main effects of competitor type, verb type, or probe point, a marginally significant interaction between competitor types and probe points for the by subject analysis only, $F1(8,232)=1.91$, $p=.059$, $\eta_p^2=.062$; $F2(8,184)=1.50$, $p=.176$, $\eta_p^2=.061$, and no significant interaction between competitor type and verb type, $F1(4,116)=2.08$, $p=.088$, $\eta_p^2=.067$; $F2(4,92)=1.89$, $p=.118$, $\eta_p^2=.076$. The analyses seem to suggest that when participants were probed at the noun, competitors that shared linguistic properties with the target object (phonological and especially semantic competitors) tended to attract attention in the initial stages of sentence processing. On the other hand, competitors that shared visual properties with the target object (shape and color competitors) tended to attract attention at the later stages of sentence processing.

As shown in the second panel of Table 6, in Experiment 1b there was a statistically significant main effect of competitor type and a marginally significant main effect of verb type for the by subject analyses only, no significant main effect of probe point, a significant interaction between competitor types and probe points for the by subject analysis only, $F1(8,224)=2.10$, $p=.037$, $\eta_p^2=.070$; $F2(8,184)=1.44$, $p=.184$, $\eta_p^2=.059$, and no significant interaction between competitor types and verb types $F1(4,112)=0.274$, $p=.889$, $\eta_p^2=.010$; $F2(4,92)=0.632$, $p=.641$, $\eta_p^2=.027$. Bonferroni post-hoc pairwise comparisons indicate that the difference between competitor types lay between the unrelated and phonological competitor objects, with greater first saccades to the unrelated pictures of objects. Therefore, when participants had a less accurate and stable internal representation and were probed at the verb a more complex picture emerged, with "linguistic" competitors attracting attention in the initial stages of sentence processing, followed by "visual" competitors, and then ending with "linguistic" competitors (especially semantic). This was, however, only significant for the by subjects analysis suggesting that which competitors grabbed participants' attention may have depended on the type of items present within each set of 6 objects.

Finally, as indicated in the third panel of Table 6, in Experiment 1c there was a statistically significant main effect of probe point, no significant main effects of verb type or probe point, no significant interaction between competitor types and probe points, $F1(8,136)=0.662$; $p= .724$, $\eta_p^2= .037$; $F2(8,184)=0.856$, $p= .555$, $\eta_p^2= .036$, and no significant interaction between competitor types and verb types, $F1(4,68)=0.830$; $p= .511$, $\eta_p^2= .047$; $F2(4,92)=0.369$, $p= .798$, $\eta_p^2= .016$. Bonferroni post-hoc pairwise comparisons indicate that the difference between competitor types lay between the phonological and semantic competitor objects, with greater first saccades to the semantic pictures of objects. Thus, when given more time to scan the visual displays participants tended to have their visual attention drawn to the semantic competitors, especially in the initial stages of sentence processing. In addition, participants in this experiment were less likely to make first saccades to the unrelated competitor items, thus, lending support to the assumption that the additional scanning time proved helpful in creating more stable and accurate internal representations of the visual displays. All of these analyses, however, need to be interpreted with caution, given the observed weak effect sizes.

1.6. Post-Test Normative Study for Competitor Objects

A possible limitation to the results discussed in the previous section is that no normative study was conducted to ensure that the set of selected pictures of objects resembled the noun referents in each of the five features (e.g., that *can-dale* and *can-dy* are seen as having the same initial phonological segment). Therefore, a similarity rating task was conducted post-test.

1.6.1 Method

1.6.2 Participants

Thirty Concordia psychology undergraduate students participated in this experiment for monetary compensation or course credit. All were native English speakers and had normal or corrected-to-normal vision.

1.6.3 Materials and Design

The experiment was conducted on a Macintosh G4 computer with a 21" color monitor (75 Hz refresh rate). As shown in Figure 24, the pictures of objects were presented as a slideshow in Microsoft PowerPoint. Although it would have been more accurate to exhaustively compare each set of 6 objects with each other (i.e., 24 sets X 20 comparison pairs), this would have required viewing 480 slides over 40 minutes. Therefore, for the purposes of time efficiency, participants only viewed each noun referent (i.e., 24 nouns) with its associated competitors (i.e., 5 competitors) to form 120 slides.

1.6.4 Procedure

After completing the consent form (see Appendix B1), a brief explanation of each feature (i.e., what is meant by shape, color, name, and function) was given. Participants had 5 seconds to view each pair of pictures, and rate similarity on a scale of 1 to 5 - whereby "1" designated "not at all similar" and "5" designated "very similar" - as to how

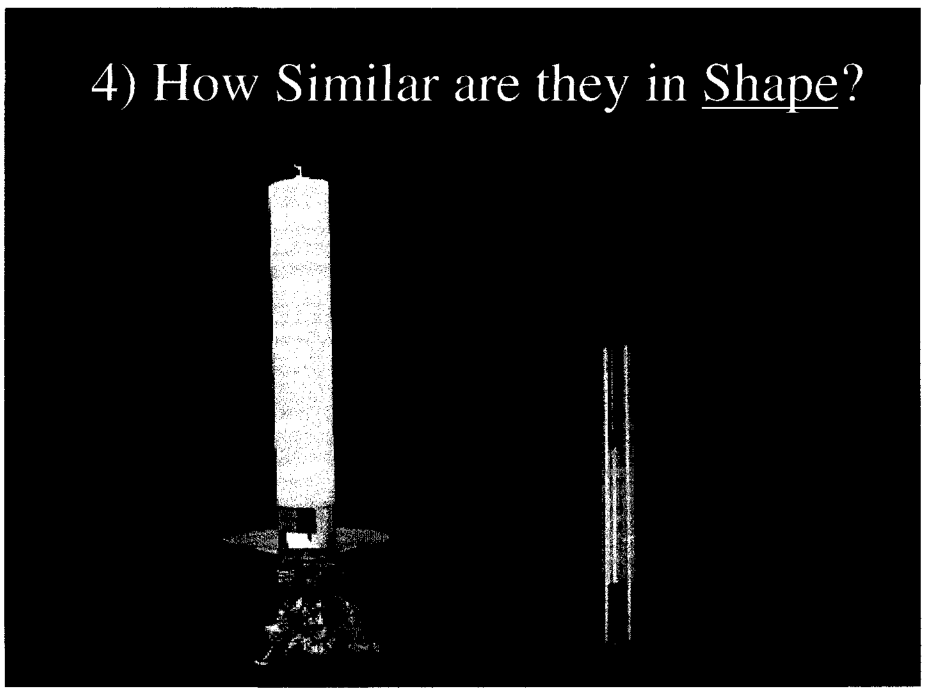


Figure 24. Sample PowerPoint slide for the competitor items normative study.

similar they were in terms of their assumed shared feature (e.g., how similar are *candle* and *pen* in shape) in a paper booklet (see Appendix A1).

1.6.5 Results

Averaging over all participants, 83.33% of the competitor items were found to be congruent with their assumed feature. The cutoff score for congruency was set to an overall rating of 2.5/5 (i.e., 50%). On average, 85.42% (i.e., 82/96) of the phonological, semantic, shape, and color competitor items received a rating over 50%, whereas 75% (i.e., 18/24) of the unrelated competitor items received an overall rating below 50%. Using the results from the normative data, incongruent items were removed and a re-analysis was conducted on the remaining items (i.e., 100/120 competitor items or 83.33% of the data).

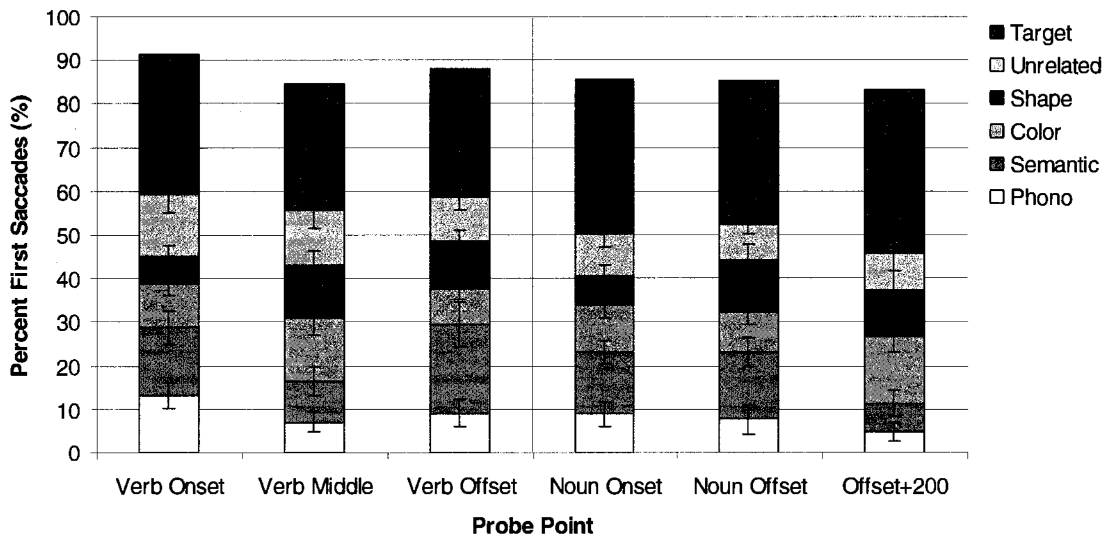
1.6.6 Competitor Objects with Incongruent Items Removed

As indicated in the bottom panels of Table 6, the analyses revealed results similar to the previous analyses that included the incongruent items. In Experiment 1a there was a statistically significant main effect of competitor type for the subject analysis only, no significant main effect of verb type or probe point, a significant interaction between competitor types and probe points for the by subject analysis only, $\underline{F1}(8,224)=2.02$, $\underline{p}=.051$, $\underline{\eta}_p^2 = 0.07$; $\underline{F2}(8,64)=1.09$, $\underline{p}=.382$, $\underline{\eta}_p^2 = 0.12$, and no significant interaction between competitor types and verb types, $\underline{F1}(4,116)=2.72$, $\underline{p}=.033$, $\underline{\eta}_p^2 = 0.09$; $\underline{F2}(4,32)=0.287$, $\underline{p}=.884$, $\underline{\eta}_p^2 = 0.035$. Bonferroni post-hoc pairwise comparisons indicate that the difference between competitor types lay between the phonological and semantic competitor objects, with greater first saccades to the semantic pictures of objects. Therefore, as illustrated in Figure 25 the same pattern of eye movements reported in the previous section emerged. Probing at the noun when participants had a poor internal representation of the visual scene appears to have initially resulted in attention being attracted by the competitors that shared linguistic properties with the target object (phonological and especially semantic competitors), and as the noun unfolded there was an increase in eye movements to the competitors that shared visual properties with the target object (shape and color competitors).

In Experiment 1b there was a statistically significant main effect of competitor type for the by subjects analysis only, a significant main effect of verb type for the by items analyses only, no significant main effect of probe point, a significant interaction between competitor types and probe points for the by subject analysis only, $F1(8,224)=2.06$, $p= .049$, $\eta_p^2 = 0.07$; $F2(8,64)=0.848$, $p= .564$, $\eta_p^2 = 0.10$, and no significant interaction between competitor types and verb types $F1(4,112)=0.265$, $p= .900$, $\eta_p^2 = .009$; $F2(4,32)=0.343$, $p= .847$, $\eta_p^2 = .041$. As demonstrated in Figure 25, when probed at the verb, it appears that "linguistic" competitors tended to attract participants' attention within the initial stages of sentence processing, followed by the "visual" competitors, and then once again ending with the "linguistic" competitors (especially semantic).

Finally, as shown in Table 6 in Experiment 1c there was a statistically significant main effect of competitor type, no significant main effect of verb type or probe point, no significant interaction between competitor types and probe points, $F1(8,136)=0.742$; $p= .655$, $\eta_p^2 = 0.042$; $F2(8,64)=1.62$, $p= .145$, $\eta_p^2 = 0.17$, and no significant interaction between competitor types and verb types, $F1(4,68)=0.684$; $p= .606$, $\eta_p^2 = 0.039$; $F2(4,32)=0.699$, $p= .599$, $\eta_p^2 = 0.08$. Bonferroni post-hoc pairwise comparisons indicate

a)



b)

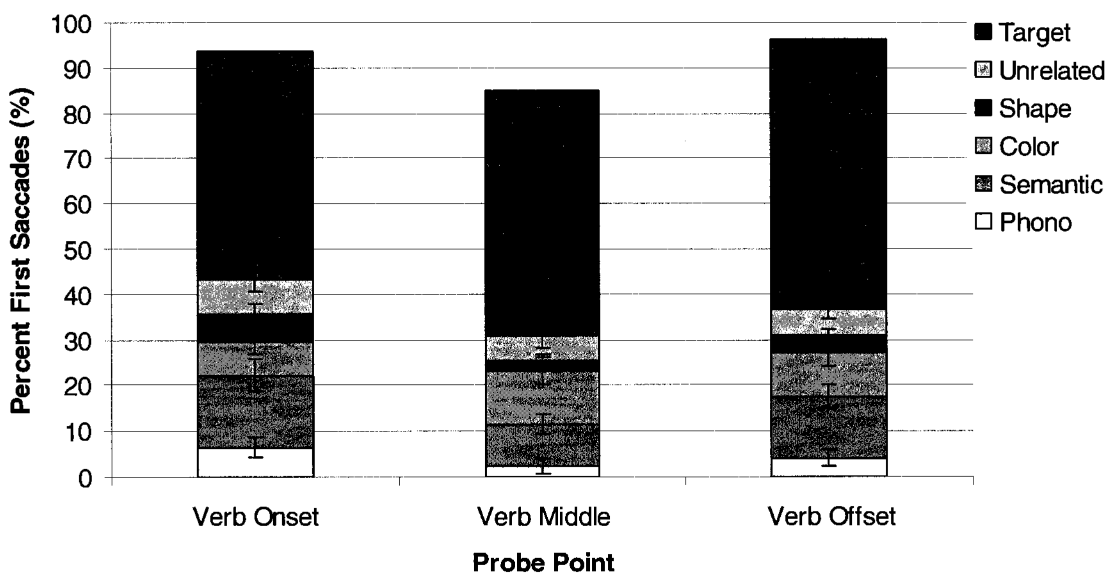


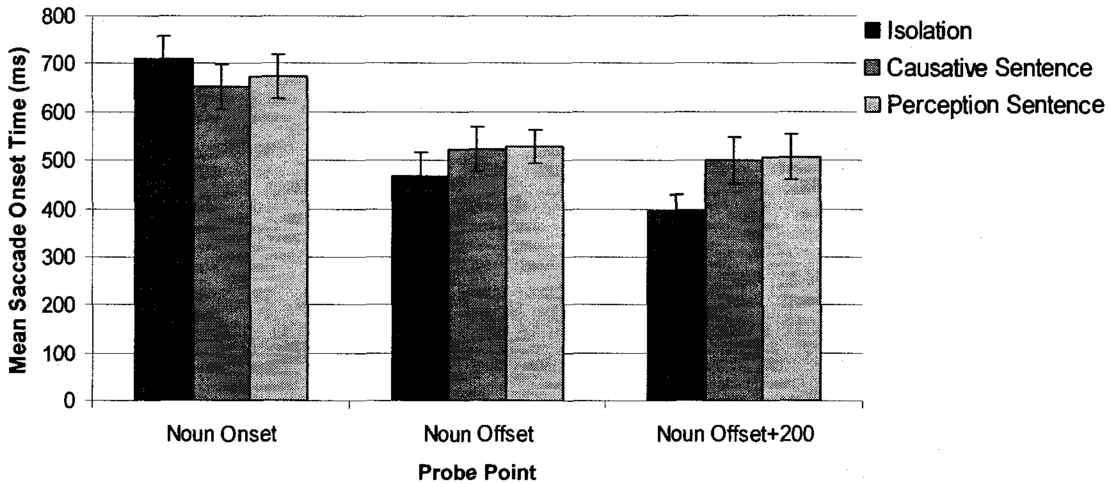
Figure 25. Percentage of first saccades made to each competitor type (with incongruent items removed), across probe points, for a) Experiment 1a and 1b, and b) Experiment 1c.

that the difference between competitor types lay between the phonological and semantic competitor objects, as well as the shape and semantic competitor objects, with greater first saccades to the semantic pictures of objects. Therefore, when participants had a more stable and accurate internal representation of the visual display, it appears that they were less likely to have visual attention attracted by the unrelated items, yet they were still drawn to the semantic competitors. The semantic competitors, although not considered the preferred direct object of the verb (according to the verb-complement study) and not mistakenly considered the verb's direct object (according to the name-matching picture task), still satisfied to various degrees the selectional restrictions of the verb (e.g., an apple can also be crushed). This factor may explain the observed delays in Experiment 1c in initiating a saccade until noun onset, the point at which unique information regarding the target object became available. Yet once again, all of these analyses need to be interpreted with caution, given the observed weak effect sizes.

1.6.7 Comparison of Current Data to Previous Research Data

A possible explanation for the pattern of results obtained in Experiment 1a emerges when they are compared to those obtained in two previous studies reported in my unpublished Master's thesis (2001). In each of these studies participants received the same set of stimuli and asked to perform the same task, but with the exception that the noun was either uttered without a sentence structure (e.g. "candle") or with no prior viewing of the visual display (i.e. similar to a visual search task). Figure 26a and 26b illustrate mean saccade onset times and percent correct first saccades respectively for the noun presented in isolation and embedded within causative and perception verb sentences (i.e., Experiment 1a), across probe points. For correct first SOTs using two one-way

a)



b)

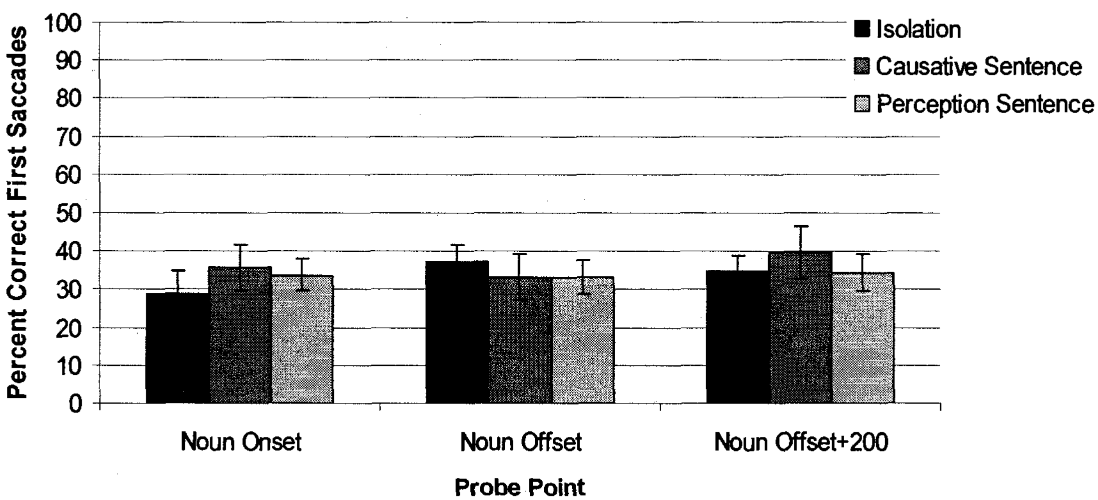


Figure 26. a) Mean saccade onset times (SOT) and b) Percent correct first saccades to target objects across noun positions, for noun presented in isolation (Van de Velde, 2001), causative and perception sentence conditions.

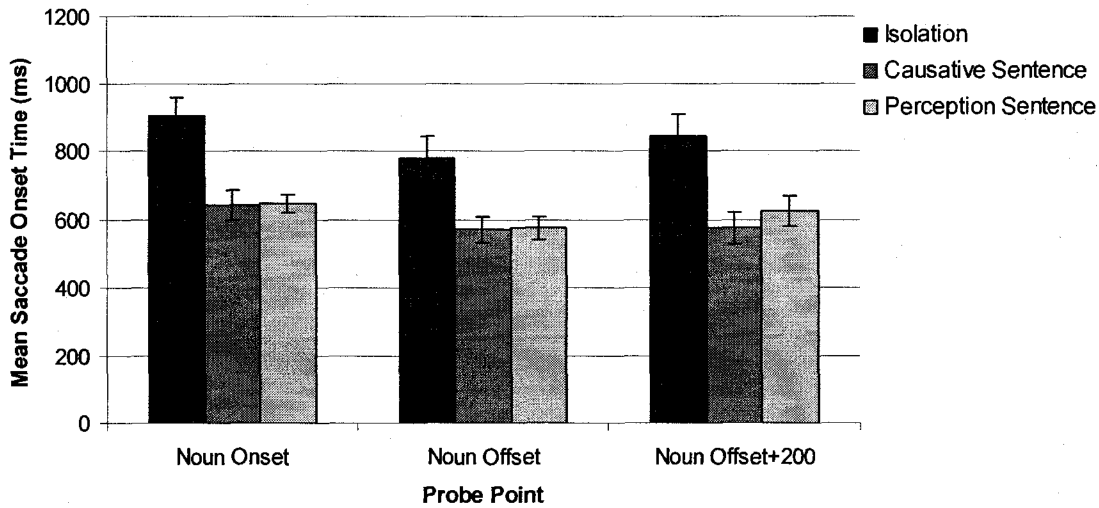
repeated measures ANOVAs, the analysis revealed no statistically significant difference across the three sentence structures (causative, perception and isolation), $\underline{F1}(2,71)= 0.355$, $\underline{p}= .702$, $\underline{\eta}^2= .010$; $\underline{F2}(2,68)= 0.201$, $\underline{p}= .819$, $\underline{\eta}^2= .006$, a significant effect of probe point, $\underline{F1}(2,142)= 20.94$, $\underline{p}= < .001$, $\underline{\eta}^2= .228$; $\underline{F2}(2,136)= 11.24$, $\underline{p}= < .001$, $\underline{\eta}^2= .142$, and no significant interaction between these two factors $\underline{F1}(4,142)= 1.06$, $\underline{p}= .380$, $\underline{\eta}^2= .029$, $\underline{F2}(4,136)= 1.15$, $\underline{p}= .335$, $\underline{\eta}^2= .033$. Thus, on average participants were not quicker at initiating a saccade to the target object when the noun was embedded within a sentence, than when it was presented in isolation. Across all three conditions, however, participants produced quicker mean SOTs as the noun naturally unfolded and then was recognized.

For percent correct first saccades using two one-way repeated measures ANOVAs, the analysis again revealed no statistically significant difference across the three sentence structures (causative, perception and isolation), $\underline{F1}(2,72)= 0.109$, $\underline{p}= .897$, $\underline{\eta}^2= .003$; $\underline{F2}(2,69)= 1.02$, $\underline{p}= .366$, $\underline{\eta}^2= .029$, no significant main effect of probe point, $\underline{F1}(2,144)= 0.317$, $\underline{p}= .729$, $\underline{\eta}^2= .004$; $\underline{F2}(2,138)= 0.891$, $\underline{p}= 0.413$, $\underline{\eta}^2= .013$, and no significant interaction between these two factors $\underline{F1}(4,144)= 0.426$, $\underline{p}= .790$, $\underline{\eta}^2= .012$, $\underline{F2}(4,138)= 1.09$, $\underline{p}= .367$, $\underline{\eta}^2= .030$. Together, these results suggest that when probed at the noun (late within sentence processing) participants did not appear to utilize the selectional restrictions of the causative verbs or the agents to constrain the domain of reference in the scene. Rather they seem to have continued to monitor the auditory input until some level of confidence was reached, that is the noun's uniqueness point.

The opposite pattern of results, however, emerged when comparing the same three sentence structures (causative, perception and isolation) in a series of studies in which there was no pre-scanning of the visual display. Figure 27a and 27b illustrate mean saccade onset times and percent correct first saccades respectively for the noun presented in isolation and within causative and perception verb sentences, across probe points. For correct first SOT using two one-way repeated measures ANOVAs, the analysis revealed a statistically significant difference across the three sentence structures (causative, perception and isolation), $F1(2,36)= 14.48$, $p< .001$, $\eta^2= .446$; $F2(2,68)= 23.51$, $p<.001$, $\eta^2= .409$, a significant effect of probe point, $F1(2,72)= 3.41$, $p= .038$, $\eta^2= .087$; $F2(2,136)= 6.42$, $p= .002$, $\eta^2= .086$, and no significant interaction between these two factors only for the by items analysis, $F1(4,72)= 0.439$, $p= .780$, $\eta^2= .024$, $F2(4,136)= 1.30$, $p= 0.274$, $\eta^2= .037$. Thus, it appears that when faced with a visual search task the visual system used all available sentential context to accelerate word recognition, before its actual uniqueness point was reached, and therefore quickened its localization within the visual display.

For percent correct first saccades using two one-way repeated measures ANOVAs, the analysis revealed no statistically significant difference across the three sentence structures (causative, perception and isolation), $F1(2,42)= 0.049$, $p= .952$, $\eta^2= .002$; $F2(2,69)= 0.034$, $p= .967$, $\eta^2= .001$, no significant main effect of probe point, $F1(2,84)= 0.764$, $p= .469$, $\eta^2= .018$; $F2(2,138)= 1.67$, $p= 0.193$, $\eta^2= .024$, and no significant interaction between these two factors $F1(4,84)= 0.370$, $p= .829$, $\eta^2= .017$, $F2(4,138)= 0.686$, $p= .603$, $\eta^2= .019$. The results, therefore, suggest that participants across all three conditions only performed at slightly above chance level (approximately 23%) since they had no prior knowledge of the location of the objects in the visual display. Thus, the sentential context may have guided participants with regards to what

a)



b)

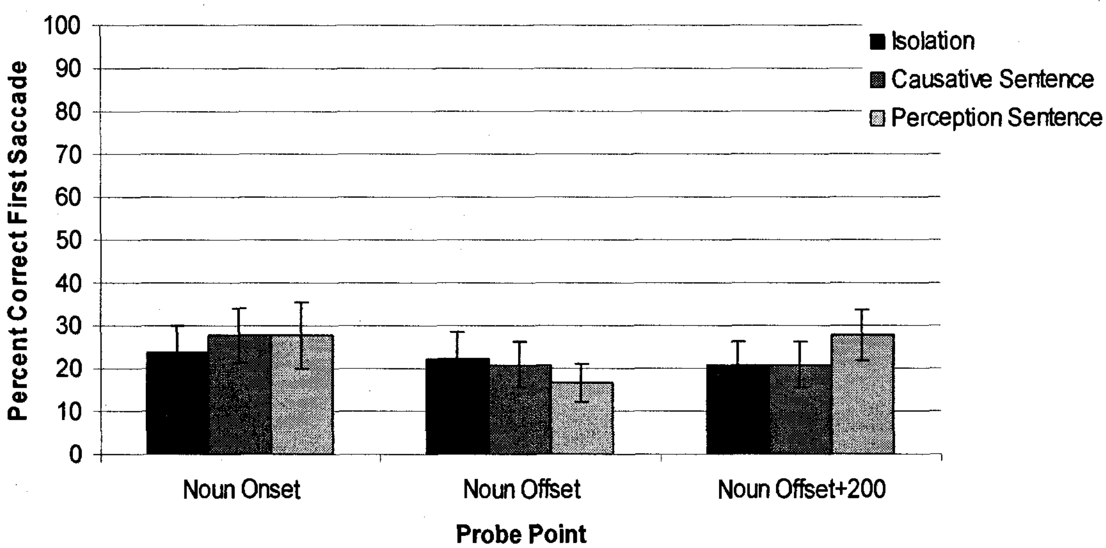


Figure 27. a) Mean saccade onset times (SOT) and b) percent correct first saccades to target objects across noun positions, for noun presented in isolation (Van de Velde, 2001), causative and perception sentence conditions, with no prior viewing of the display.

object to look for, but because they had no prior viewing of the display participants did not know where that object would be located.

Furthermore, as expected, when given the opportunity to pre-scan the visual display participants were on average quicker to initiate a saccade to the target object, $F1(1,22)= 24.34$, $p< .001$, $\eta^2= .525$; $F2(1,46)= 52.7$, $p< .001$, $\eta^2= .534$, and made fewer first saccade errors, $F1(1,22)= 4.12$, $p= .055$, $\eta^2= .158$; $F2(1,46)= 7.27$, $p= .010$, $\eta^2= .136$, when the noun was presented in isolation, than when they saw the objects for the first time simultaneously with the linguistic input (i.e., compare Figure 26 with Figure 27). Similarly, when the noun was embedded within a sentence participants were on average quicker to initiate a saccade to the target object, $F1(1,46)= 7.80$, $p= .008$, $\eta^2= .145$, $F2(1,44)= 1.40$, $p= .243$, $\eta^2= .031$, and made less first saccade errors $F1(1,46)= 7.80$, $p= .008$, $\eta^2= .145$, $F2(1,46)= 5.65$, $p= .022$, $\eta^2= .109$, when they were allowed to pre-scan the visual display.

1.7 Discussion

Participants' task in all of these studies were twofold, (1) respond as quickly as possible to whether the object referred to in the sentence was present in the visual display, and (2) if it was, make a fixation to the target object before responding on the button box (these two events usually occurred concurrently). In order to accomplish these tasks it appears that participants employed different strategies depending on the nature of their internal representation of the visual scene and the position within the sentence at which they were probed.

When participants were probed at the noun (Experiment 1a), the more semantically restrictive causative verbs did not result in fewer first saccade errors or quicker SOTs to the target object, as compared to the less semantically restrictive perception verbs. In addition, participants did not make fewer first saccade errors or initiate saccades to the target more quickly when it was embedded within a sentence structure, than when presented in isolation. These results do not appear to reflect a decaying internal representation of the visual display as the sentence unfolded or the consequence of having to inhibit saccades until the utterance of the noun, because participants made fewer first saccade errors and were quicker to initiate a saccade to the target objects, as compared to when they were *not* given the opportunity to pre-scan the visual display before the onset of the linguistic input (i.e., visual search task). Moreover, the results do not seem to reflect decay of the activated conceptual representations, because participants were able to use the available sentential context to guide visual attention to the target objects when given the same stimuli in the visual search task.

Instead the pattern of results seems to point towards a task-specific strategy. It is possible that by the onset of the noun, information extracted from the more semantically restrictive verb isolated a noun candidate. But rather than risk an incorrect saccade in a time-pressured task, participants may have opted to continue monitoring the sensory input until the noun's uniqueness point was reached, and thus the target object confirmed. Such a strategy, however, would most likely not have proven efficient in the visual search task, in which participants did not have prior knowledge of the identity or location of the objects within the visual display. Rather than wait until the noun became uniquely identifiable before starting the visual search, participants appear to have used the sentential context to initially guide visual attention to possible target objects.

The cautious strategy of continuing to monitor the linguistic input until a sufficient level of confidence was reached (e.g., uniqueness point), seems warranted in view of the presence of phonological and semantic competitors, which could be mistaken for the target object. As apparent by the high number of first saccade errors (~ 66%), approximately 12.5% of which were made to the conceptually plausible semantic candidates and 9.1% to objects which shared similar sound onsets. Moreover, this is in line with previous findings by Tanenhaus and colleagues (Tanenhaus et al., 1995; Allopenna, Magnuson & Tanenhaus, 1998; Dahan, Magnuson, Tanenhaus & Hogan, 2001; Dahan, Magnuson & Tanenhaus, 2001), who demonstrated that the presence of competitor objects delayed eye-movement latencies to the target objects. Based on Marslen-Wilson's (1989, 1990) cohort model, they argued that the delays in launching the appropriate saccades occurred as a result of the lexical and semantic candidates becoming sufficiently activated to compete for recognition with the target noun. Moreover, recent studies have shown that when passively listening to a sentence, participants were more likely to initially fixate on competitor objects that share the same conceptual category as the target object (e.g., "piano" and "trumpet") (Huettig & Altmann, 2005; Yee & Sedivy, 2006).

A different story emerged, however, when participants were probed at the verb and given a short time to scan the visual display (Experiment 1b). The large number of first saccade errors suggests that 6 seconds to scan the visual display, as compared to 24 seconds in Experiment 1c, was insufficient time to elicit a stable and accurate internal representation of the scene (i.e., memory trace for the identity and location of all 6 objects on the screen). This poor internal representation appears to have resulted in participants using the semantically restrictive nature of the causative verbs to produce anticipatory eye movements to the most probable direct objects of the verbs. This verb advantage persisted even when the confounding variable, verb length, was accounted for. Saccades were launched before the onset of the noun in 27.0%, 21.6% and 9.77% of the causative trials, compared to 28.7%, 11.2% and 5.17% of the perception trials, for the verb onset, middle and offset conditions respectively. Thus, after correcting for some of the methodological problems and when adding competitor objects in the visual display, fewer anticipatory eye movements to the target objects were obtained than in Altmann and Kamide's (1999) study – 32% of the semantically restrictive trials and 18% of the semantically non-restrictive trials.

Moreover, in Experiment 1b saccades were launched towards the target object in the causative trials on average 32 ms, 78 ms and 321 ms after the onset of the noun for the verb onset, middle and offset conditions, respectively, as compared to 156 ms, 211 ms, and 269 ms in the perception trials. If we take the programming of a saccade to necessitate approximately 200 ms (Matin et al., 1993), then the saccades for the verb onset and middle conditions were actually initiated during the lifetime of the verb. Once again, these values were similar to those reported by Altmann and Kamide (1999), with first saccades occurring 127 ms after the onset of the noun in the semantically nonrestrictive condition and 85 ms *before* the onset of the noun in the semantically restrictive condition. It is important to note, however, that first saccade latencies to the target object relative to noun onset did not remain constant or decrease, but rather increased as the verb unfolded. This suggests that initially, during the utterance of the verb, visual attention was automatically guided by the semantic restrictions of the causative verb. But then by verb offset, with the syntactic knowledge that the target would be uttered next, it appears that participants employed a similar strategy as in Experiment 1a – to delay saccades until after noun onset.

Yet, once again, a different pattern of results emerged in Experiment 1c, in which participants were given more time to pre-scan the visual display. As demonstrated by the significant reduction in first saccade errors, when given more time to study the visual display participants appear to have formed a more stable and accurate internal representation of the visual array of objects. These more accurate internal representations, however, did not result in quicker mean SOTs. In fact, correct first SOTs were significantly slower, as compared to eye movement performance in Experiment 1b. Participants in this context appear to have been more cautious (i.e., substituting accuracy for speed), producing fewer anticipatory eye movements, and instead waiting until the target nouns were verbally encountered before initiating their first saccades. Saccades were launched towards the target objects before the onset of the noun in 20.4%, 17.1% and 5.55% of the causative trials, compared to 17.1%, 11.1% and 2.78% of the perception trials, for the verb onset, middle and offset conditions, respectively. As in Experiment 1b, however, participants were quicker to initiate a saccade to the verb's direct object when it was preceded by the causative verb, even after accounting for differences in verb length. The more semantically restrictive nature of the causative verbs, thus, appears to have aided in accelerating the noun's recognition point as it was unfolding in the sentence.

Two possible explanations for the delayed saccade onset times when given 24 seconds to pre-scan the visual display, are that: (1) participants had ample time to name each object as a memory strategy, which they may have then attempted to match against the unfolding target noun; and/or (2) the occurrence of a more in depth conceptual representation of the visual objects may have resulted in the activation of objects that satisfied any of the selectional restrictions of the verb to compete with the target noun for recognition. No evidence is currently available to support the former explanation, as participants were not asked after Experiment 1c whether they had consciously named each object in the display. There is, however, some evidence for the latter explanation, as indicated by the higher proportion of first saccade errors to the conceptually plausible semantic candidates, as compared to the other competitor objects. In the two proposed scenarios participants would be required to wait for the sentences to unfold until the target nouns became uniquely recognizable, before being able to initiate their saccades with confidence. This recognition point occurring sooner for the causative trials, as its semantic restrictions would be more helpful at initially constraining the domain of reference in the scene.

Finally, an examination of the pattern of eye movements seems to indicate that during sentence processing, objects that shared certain physical features with the target also became activated and thus competed for attentional resources. Early within the processing of the verb-noun complement, objects that were related to the target along both phonological and semantic dimensions appear to have become activated. By contrast, during the later stages of processing, stored information about visual form and color appear to have become activated and possibly used to guide visual attention towards objects in the display that shared the same visual features as the target object (e.g., "candle" = search for something thin, cylindrical and white). Although the obtained effect sizes were small and normative measures were only conducted post-hoc, Huettig and Altmann (2004) have reported similar findings in a series of experiments. The authors demonstrated that language-mediated visual attention was immediately directed towards objects that shared similar perceptual (color and form) and semantic properties with the target word. Huettig and Altmann (2004), however, did not examine the influence of multiple competitor types when present concurrently, and did not provide a natural time-course of their activation during sentence processing. Further research using the visual world paradigm will thus be required to understand in greater detail what information during sentence comprehension is activated and when.

In summary, the processor appears to incrementally integrate all available knowledge (linguistic and non-linguistic) with the aim to rapidly interpret linguistic descriptions of what we see in the world and how we may interact with it. In particular, the results from this series of experiments provide additional confirmation for the notion that syntactic and semantic information extracted at the verb can be used to constrain the domain of reference in the scene, and in some cases predict the referent of the grammatical complement of the verb, *before* it is uttered in the sentence (Altmann and Kamide, 1999; Kamide et al., 2003). These anticipatory eye movements, however, do not appear to be mandatory or generalizable across all situations. Rather its rate of occurrence seems to vary as a function of the strength of the memory trace and conceptual representations of the objects in the visual scene, the presence of concurrent competitor objects, as well as task demands. Although these results clearly show that visual context affects sentence processing, the locus of the interaction between these two systems still remains unclear. The obtained results can, on the one hand, be interpreted as evidence that the *early* processes of the linguistic system are not modular, but rather influenced by information processed by the visual system and higher cognitive systems. On the other hand, the results fail to reject the hypothesis that visual context affects processing only *after* the linguistic system has initially analyzed the sentence, to select alternative sentence parsings or interpretations for further processing or re-analysis.

Limitations to this Study

In this series of experiments the visual displays consisted of a group of objects arrayed in a circular pattern. This type of visual display, which Henderson and Ferreira (2004) refer to as "Ersatz scenes" (German word meaning substitute, which in English came to connote something inferior or phony), are commonly used by experimenters in psycholinguistics using the visual world paradigm (Tanenhaus, Altmann, and their colleagues). Although arrays of objects (real or depicted) allow for greater experimenter control, as compared to true scenes, Henderson and Ferreira (2004) proposed that the two types of visual displays differ significantly in how they are processed visually and cognitively. First, in true scenes there is a coherent meaningful relationship between the objects in the scene, that allows for rapid identification of a scene category (e.g., an office) and the subsequent retrieval of additional semantic information associated with that category (e.g., usually contains a desk, a chair and a computer). This is not possible with arrays of objects that contain no natural structure and, therefore, no higher level meaning (i.e., scene gist) beyond that provided by the individual objects. This point is particularly important when the aim of the study is to determine how semantic information extracted from the visual context affects language processing. Second, objects in arrays tend to violate proportionality properties (e.g., a ball is usually not 3 times the size of a boy's head), and spatial licensing constraints (e.g., lit birthday cakes are not typically found on the floor). Furthermore, the absence of background detail in arrays often results in larger objects, as the scenes are less dense and objects are not occluded. These major differences, although only two of many more, raise questions as to whether the results obtained with arrays, such as the present study and all studies up to date, generalize to real-world scenes.

Another concern is that the instructions given to participants in this series of studies may have resulted in eye movement behavior that was not representative of actual every day language processing. The task demand, to respond as quickly as possible to whether the object referred to in the sentence was present in the visual display, may have prompted participants to "artificially" anticipate the identity of the target noun before it was uttered in the sentence. Evidence for this hypothesis can be seen in the high number of first saccade errors, the presence of anticipatory eye movements in the perception conditions (particularly early within sentence processing), and the shorter SOTs towards incorrect objects. In addition, the judgment task placed additional attentional and processing demands on participants, which may have interfered with language processing and, therefore, question the ecological validity of the study. These concerns were in fact addressed in the second part of Altmann and Kamide's study (1999) and demonstrated that when participants were not given a judgment task anticipatory eye movements were significantly reduced. One can argue, however, that the instructions:

> "Each picture will be accompanied by a short sentence spoken over the loudspeakers, but in this version of the experiment we aren't asking you to pay any particular attention to the sentences (some refer to the things in the pictures, and others don't, but that isn't relevant to this experiment)" (Altmann and

Kamide, 1999, p.256)

may have still provoked participants to develop a specific task strategy, because they were either able to correctly deduce that deception was being employed by the experimenter or were victim of the “reverse psychology” phenomenon (e.g., “Don’t look down!”).

These inherent problems, as well as a lack of studies using true scenes, prompted my fellow colleagues (Di Nardo, 2005; de Almeida, Di Nardo & von Grünau, in preparation) to examine how the linguistic and visual systems interact in the context of naturalistic photographs and dynamic scenes. In a series of experiments de Almeida et al. (in preparation) presented participants with still film clips or movies in which the agent was seen as moving (apparent or actual) either towards or away from the target object, or neither (See Figure 28). Similar to the study presented in this paper, they contrasted the more semantically restrictive causative verbs against the less semantically restrictive perception/psychological verbs. Participants were not given any specific task, but were simply asked to look at the stills or movies and pay attention to the sentence. At the end of the experiment, however, participants were given a short recognition task to ensure that they had been paying attention. Eye movement data indicated that saccades were initiated sooner in the “towards” condition (stills: 170 ms after noun offset; movies: 145 ms after) than in the “away” (stills: 552 ms after; movies: 492 ms after) and “neutral” (stills: 626 ms after; movies: 717 ms after) conditions. Moreover, verb effects were only found in the “towards” condition, with faster SOTs in the causative condition than in the perception condition. Finally, there were few anticipatory eye movements to the object referent of the grammatical complement of the verb; occurring in only 6.3% of causative

a)

b)

c)

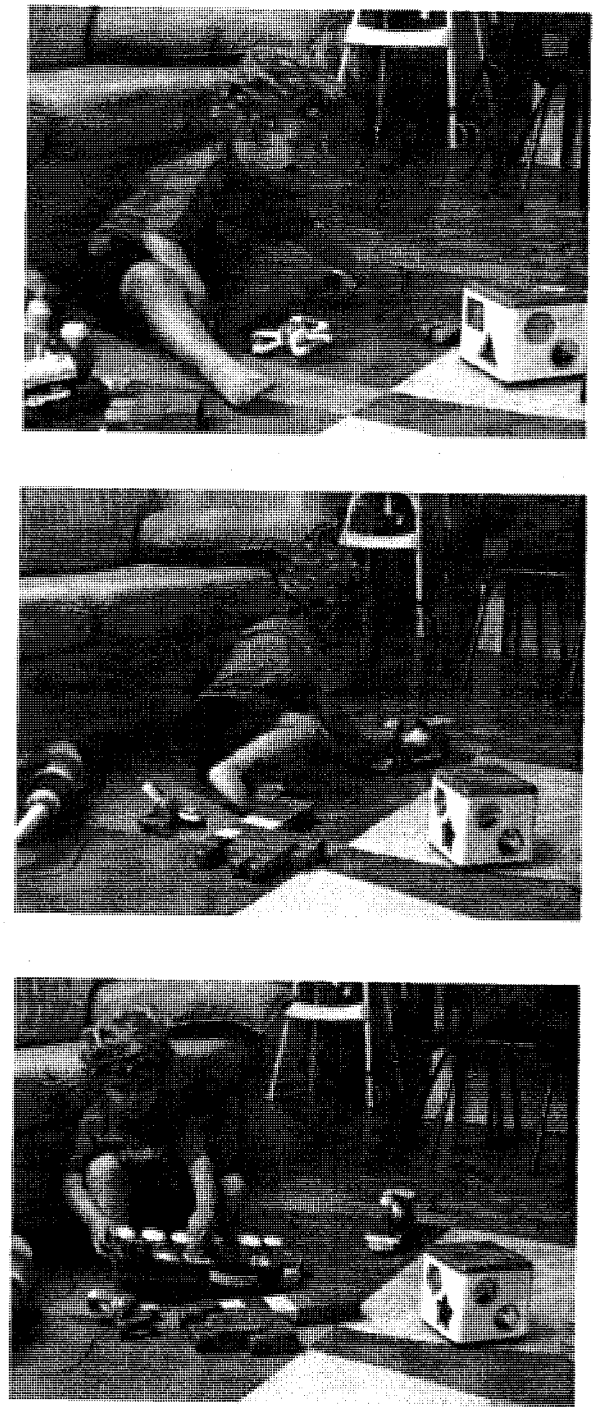


Figure 28. Sample frames from a dynamic scene, which were accompanied by either of two sentences: 1) “While playing with the toys, the boy will *roll* the cube” (causative event); 2) “While playing with the toys, the boy will *check* the cube” (perception/psychological event). The three frames represent the three motion conditions (a) Towards; b): Away; and c) Neutral) of the agent of the event (the boy) with respect to the target object of the sentence (the toy cube) (de Almeida, Di Nardo, & von Grünau, in preparation).

trials and 9.7% of perception trials with the still photographs, and 7.5% of causative trials and 9.3% of perception trials with the movies.

De Almeida et al. (in preparation), concluded that their findings fail to support the notion that the linguistic system plays a major role in guiding eye movements in the early stages of linguistic and visual interaction. Rather the results suggest that when using true scenes, in particular movies that depict everyday events, visual context takes precedence over the properties of the linguistic input in constraining the domain of subsequent reference. Only when the verb-specific information was consistent with the visual context (i.e., the motion of the agent), were saccades launched more quickly to the target object, which the authors argue represented a confirmatory effect, rather than an anticipatory one. That is, information extracted from the sentence was used to confirm expectations about the likely events that were anticipated from the scene gist. On the other hand, when visual and linguistic context provided inconsistent information (i.e., "away" condition) additional time may have been required by the central processor to integrate the two inputs. Finally, de Almeida et al. (in preparation) surmised that when the visual input is relatively impoverished (i.e., lacking scene gist, naturalistic agents, events, and motion), as is the case with arrays of objects and pictures of non-complex scenes, that only then does the attentional system appear to be controlled by the ongoing process of the linguistic input.

Rationale for the Second Series of Experiments

The findings by de Almeida et al. (in preparation), that verb-specific information did not serve to constrain domains of visual reference, were in contradiction to the work conducted by other researchers within the visual world paradigm (Tanenhaus, Altmann and their colleagues). The use of naturalistic scenes, as opposed to arrays of objects and pictures of non-complex scenes, appears to be the chief culprit for these incongruent findings. As de Almeida et al. (in preparation) were the first to employ dynamic naturalistic scenes to study the influence of visual context on language comprehension, the purpose of the second series of the present studies was to extend these findings but with the use of different classes of verbs, denominal and non-denominal verbs, which also allowed us to investigate how much information about a scene viewers can encode.

According to Clark and Clark (1979) a noun can be used as a verb (usually nouns that name objects), called “denominal verbs”, if the speaker believes that the listener can readily infer the meaning of a verb given their mutual knowledge. For instance, when a speaker utters the sentence “The woman will vacuum the carpet”, he/she assumes that the listener will understand that the noun “vacuum” acts as both the verb and the instrument that will be used to complete the action. On the other hand, when a speaker uses non-denominal verbs such as “The man will carve the turkey”, he/she is required to add an instrumental case (i.e., "with the use of", "because of", or "by means of") to the noun to denote what instrument the action will be performed with (e.g., “with the knife”). Thus, similar to the first set of studies two verb classes were contrasted, which also differed in their semantic restrictiveness. Although non-denominal verbs (e.g., chop) are semantically restrictive, in that their grammatical objects (e.g., vegetables) and instruments (e.g., knife) are constrained by the properties of their referents (e.g., chop-able object, and instrument capable of chopping the object), they are less restrictive than denominal verbs (e.g., to vacuum), for which the instrument noun is invariably identical to the verb (i.e., vacuum). Thus, of interest was to understand how the verb’s selectional restrictions, and possibly those of the verb's preferred grammatical complement, direct attention (and the accompanying eye movements) toward the referents of the object and instrument nouns during sentence comprehension. While also investigating how visual representations may influence the selection of conceptually plausible grammatical objects and instruments of verbs.

Previous studies using the visual world paradigm seem to have neglected two important factors in the control of eye movements: (1) that covert attentional shifts can be made without overt eye movements; and (2) that participants may not encode rich information about a visual scene and its constituent objects early on during scene processing. Our ability to allocate attention to different parts of the visual field while maintaining fixation has long been documented (Posner, 1980). This poses an important limitation in eye tracking research, because the resulting scan patterns will not necessarily show where a participant's attention has been and, therefore, will not fully reflect underlying cognitive processing. Furthermore, the notion that participants encode and retain detailed visual representations may be justified with the use of simple displays, but seems unlikely with complex dynamic displays. These naturalistic displays often contain vast amounts of detail, as well as agents that naturally attract an observer's attentional system. It is quite possible, therefore, that the contradictory findings obtained by my colleagues (de Almeida et al., in preparation) are, in part, caused by the visual system's limited capacity to encode and/or retain visual details into memory. Instead, when faced with complex dynamic scenes, observers may have traded visual detail for more abstract information, to form a representation of the scene category or gist. Consequently, the delayed SOTs and small number of anticipatory eye movements obtained by de Almeida et al. (in preparation) may represent a failure of the visual system to attend and encode the target objects before the onset of the verb.

In the last 10 years, a group of researchers has extensively studied the circumstances in which our visual system fails to notice large changes to scenes (now referred to as "change blindness") to gain insight into the nature of our internal representations about the visual world (e.g., Grimes, 1996; Simons, 1996; Levin & Simons, 1997; Rensink, O'Regan, & Clark, 1997; O'Regan, Rensink, & Clark, 1999; Simons & Levin, 1998; Simons, Franconeri & Reimer, 2000; Hollingworth, Schrock & Henderson, 2001; Hollingworth & Henderson, 2002; Henderson & Hollingworth, 2003; Hollingworth, 2003). Real-world examples of the change/inattention blindness phenomenon abound; and we can all recall instances in which we have been fooled by a magician's illusion, failed to notice a friend in a crowded bus, or failed to perceive our colleague's new haircut. Thus, before going any further, pertinent findings in the area of change blindness research will be reviewed, as they provide an important backdrop to the second set of studies.

Change Blindness

The study of instances in which misperception occurs in order to gain insight into how perception normally operates is quite a common strategy (e.g., optical illusions), and is not restricted to the area of vision research. For example, researchers frequently study mental illness and brain damage/disorders to understand normal functioning of the mind and brain. The phenomenon of change blindness has become a popularized instance of when the visual system breaks down, that is when it fails to detect dramatic changes in scenes, particularly when the change occurs during a visual disturbance. Although, change blindness had been previously reported (e.g., Phillips, 1974; Bridgeman, Hendry & Stark, 1975; McConkie & Zola, 1979; Pashler, 1988; McConkie & Currie, 1996), Grimes (1996) was the first to examine this phenomenon with the use of natural scenes. In his study, while participants were examining photographs of natural scenes for a future memory task, changes to the image were synchronized with the advent of a saccade. Participants were found to be surprisingly poor at detecting large changes in the scenes (e.g., only 50% noticed two people exchanging heads); changes that were easily noticeable when they occurred during a fixation. Grimes (1996) hypothesized that the smear of the visual input on the retina, caused by the rapid movement of the eye, served to mask the local transients (i.e., visual disturbances) that normally arise with sudden changes in scenes. The findings by Grimes (1996) of poor change detection were also in line with earlier work, which demonstrated that the visual system is quite limited in its capacity to encode, retain, and compare visual information across saccades (e.g., Rayner & Pollatsek, 1983, 1992; Irwin, Yantis & Jonides, 1983; Jonides, Irwin & Yantis, 1983; Pollatsek, Rayner & Collins, 1984; Irwin, 1991).

The study by Grimes (1996) created a renewed interest in the study of change blindness, and led to the discovery that similar failures could occur in the absence of eye movements (i.e., during a fixation). One such instance is with the use of the "flicker paradigm", which as illustrated in Figure 29 consists of displaying in rapid alternation an original and modified scene interspersed with a blank screen (giving the display a flickering appearance) (e.g., Rensink et al., 1997, 2000; Simons, 1996; Rensink, 2004). This global flicker serves to mask the local transients caused by the change and, therefore, prevent attention from being captured by the location of the changed object.

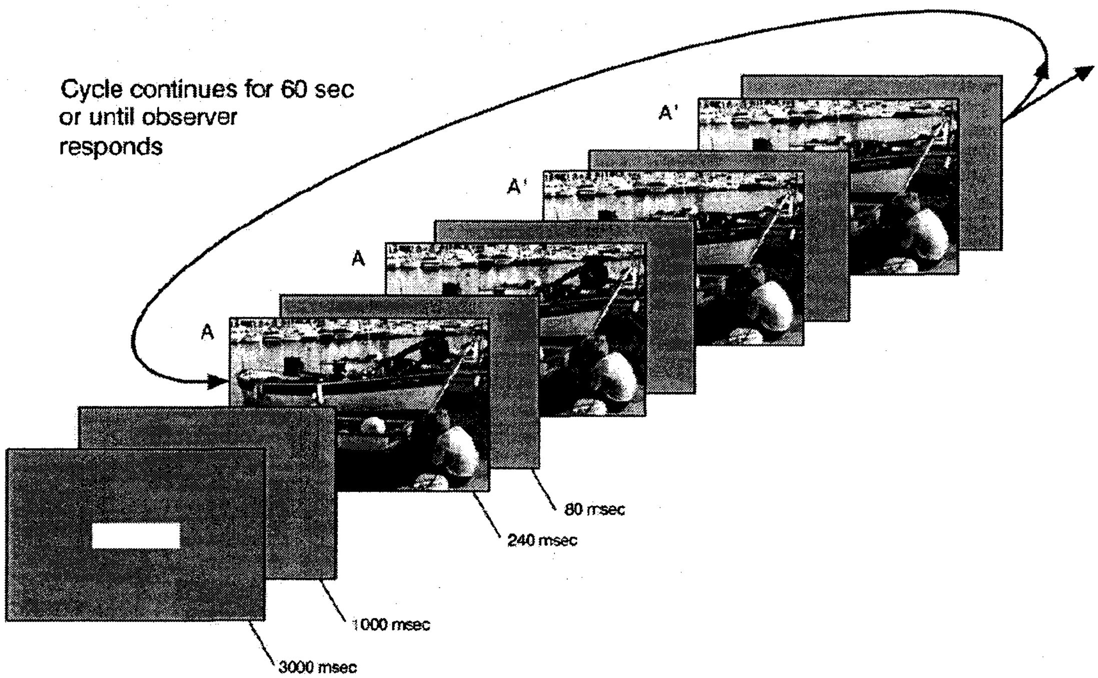


Figure 29. Sequence of events in the flicker paradigm from Rensink, O'Regan et al. (2000) study, p. 129. Each trial began with a 3 second grey screen containing a white rectangle to alert the participant that the trial had begun. This was then followed by a 1 second grey screen. The original image (A) was then displayed for 240 msec, and followed by a blank screen for 80 msec. After two cycles, the modified image (A′) was then shown for 240 msec, followed by the blank screen for 80 msec., and repeated for another cycle. This alternating sequence continued until the participant responded or 60 seconds had elapsed.

Although in Rensink et al.'s (1997) study participants were armed with the knowledge that a scene change would occur (i.e., an intentional task), they were surprisingly slow at detecting the change (17.1 alternations or 10.9 seconds). However, participants were found to be twice as fast at identifying changes to central interest areas (i.e., mentioned by at least 3/5 observers in an independent study), as compared to marginal interest areas (i.e., not mentioned by any of the observers). Overall, these results suggest that the phenomenon of change blindness can be obtained without saccades and, therefore, is not related to a saccade-specific suppression mechanism. Rensink et al. (1997) concluded that detection requires focused attention to the changing item during and after the change. Whereas "in the absence of focused attention, the contents of visual memory are simply overwritten (i.e., replaced) by subsequent stimuli, and so cannot be used to make comparisons" (Rensink, et al., 1997).

Rensink (2000a, b, 2002) later proposed a theory, the coherence theory, to describe how attention is allocated in scene perception. The theory suggests that while viewing a scene we rapidly form volatile proto-objects (i.e., units of visual information), only a few of which are then selected by focused attention to build a coherent and stable object representation. It is only during this time, when the object is given focused attention, that a change can be detected. However, because attentional resources are limited to only one or few objects at any one time, initially many changes are likely to go undetected – particularly when attention are not captured by the low-level signals that accompany the change. Furthermore, according to the coherence theory, after focused attention is released, the object loses its coherence and reverts back into its volatile constituent proto-objects. Thus, information within the visual short-term memory (vSTM) is quickly overwritten. Finally, Rensink (2000a,b, 2002) proposes that perceptions of scene gist and layout (retained in short-term representations) and scene schemas (retained in long-term representations), provide observers with stable scene constraints, such as which objects to expect, their likely location, and their importance for the task at hand, to carefully co-ordinate attentional shifts and eye movements. Other factors which also appear to influence the orientation of attentional focus include semantic information (e.g., Hollingworth & Henderson, 2000), individual differences (e.g., Werner & Thies, 2000), and task demands (e.g., Wallis and& Bülthoff, 2000).

Concerns that in the "flicker" paradigm processing at the location of the change may have been interrupted or erased by the masking effects of the blank screens or a disruption in visual continuity (due to the flicker), led O'Regan et al. (1999) to create the "mudsplash" paradigm. As shown in the top panel of Figure 30, six small high-contrast rectangles were dispersed over the picture at the moment of the change, giving the impression of mud being splashed on a car windshield. O'Regan et al. (1999) argued that because the "mudsplashes" provoked only minor visual disturbances and never covered the location of the change, an occurrence of change blindness could not be attributed to masking or erasure of the internal representation. Each pair of pictures cycled until participants pressed a button to indicate that they had perceived the changed area or object, which could take the form of a change in location, color, appearance or disappearance. Results indicated that central-interest changes were usually detected within the first cycle, whereas marginal-interest changes required two or more cycles. In fact, for 13–30% of the trials participants failed to notice the change after 40 seconds of viewing. First, these results demonstrated that global masking was not necessary to

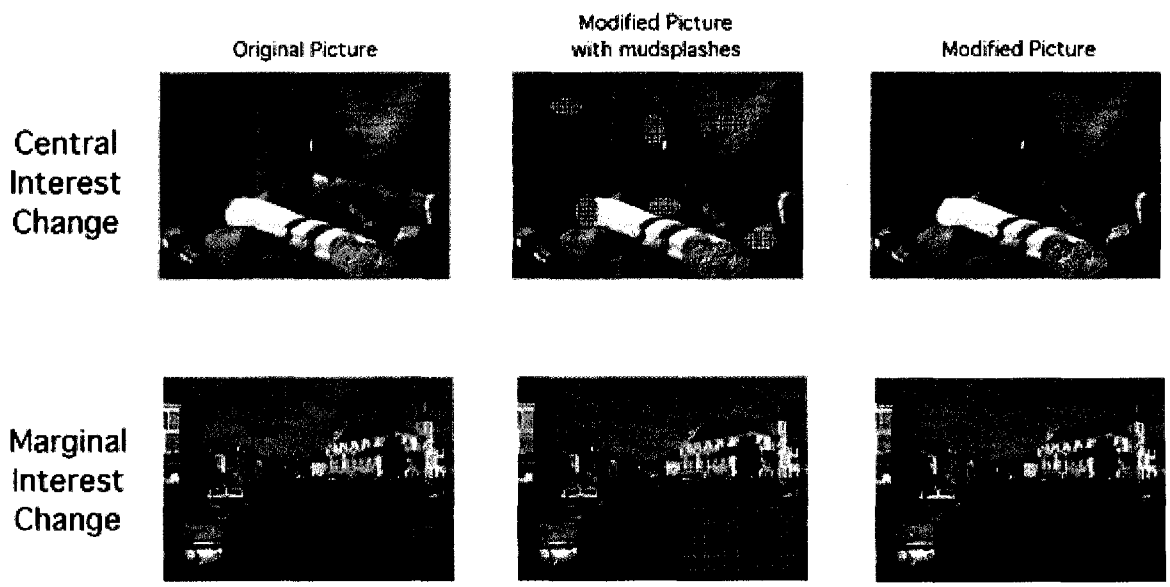


Figure 30. Sequence of events for the "mudsplash" (top panel) and masking rectangle (bottom panel) experiments, with changes occurring to either central or marginal interest areas from O'Regan et al. (1999) study (supplemental information).

produce change blindness, but could also occur when brief extraneous transients diverted attention from the location of the change signal. More importantly, however, the results suggest that observers mainly attend and encode central-interest information from their visual world, internal representations of which are then available for making comparisons across views.

Further evidence for this last interpretation was provided in their second experiment (O'Regan et al., 1999). During the moment of the change, instead of "mudsplashes" a single black-and-white textured rectangle covered the location of the change (see bottom panel of Figure 30). Although participants were cued as to the location of the change, they were significantly more accurate at identifying central-interest changes than marginal-interest changes. These results, thus, indicate that the insertion of a mask which covered the location of the changed object or area did not erase its mental representation. But most importantly, the results demonstrated that the observers had only encoded the central-interest items, which were then available for comparison. O'Regan et al. (1999) concluded that a rich detailed representation of the entire scene is not necessary because the visual world acts as an external memory, available for immediate reference or further inspection with a quick movement of the eye (also see O'Regan, 1992).

Although these studies provide evidence for the notion that focused attention is necessary to detect central-interest changes, previous findings (Simons, 1996; Ballard, Hayhow & Pelz, 1995) in which change blindness occurred for attended objects, suggest that focused attention is not sufficient. Simons and Levin (1997a, b) hypothesized that successful change detection also requires effortful processing of the attended objects. To test their assumption, Simons and Levin (1997a, b) asked participants to pay close attention to a short video depicting a single actor performing a simple action (e.g., walking to a ringing phone and answering it). As shown in Figure 31, when the camera cut to a different view, the main actor was switched. After viewing the short clip, participants were required to write a brief description of the video they just saw. If no change was reported, then participants were asked directly if they had noticed the actor change. Analysis of the written responses indicated that although they were able to provide a detailed description of the actor's clothing, surroundings and motion, the center of their attention, only 33% of the 40 participants noticed the actor change. This poor performance was in sharp contrast to the near perfect performances obtained in the subsequent study, in which prior to viewing the movie, participants were given a schematic illustration of each shot in the video, as well as information regarding when during the video an actor change could occur (on 50% of the trials). Simons and Levin (1997a, b) concluded that change detection is not an automatic process, but rather requires effort on the part of the observer. In order to form a rich detailed representation of the visual world that can be preserved across different views of a scene, observers need to deliberately encode the visual properties of all attended objects.

Although Simons and Levin's (1997) use of dynamic scenes significantly increased the ecological validity of their change blindness study, an important limitation remained. Their participants were not studied in real-world situations where they could actively interact with the objects or people in their environment. The fact that they were only passive observers of their visual world may have engendered a less detailed representation of the scene. Thus, in an innovative study, Simons and Levin (1998)

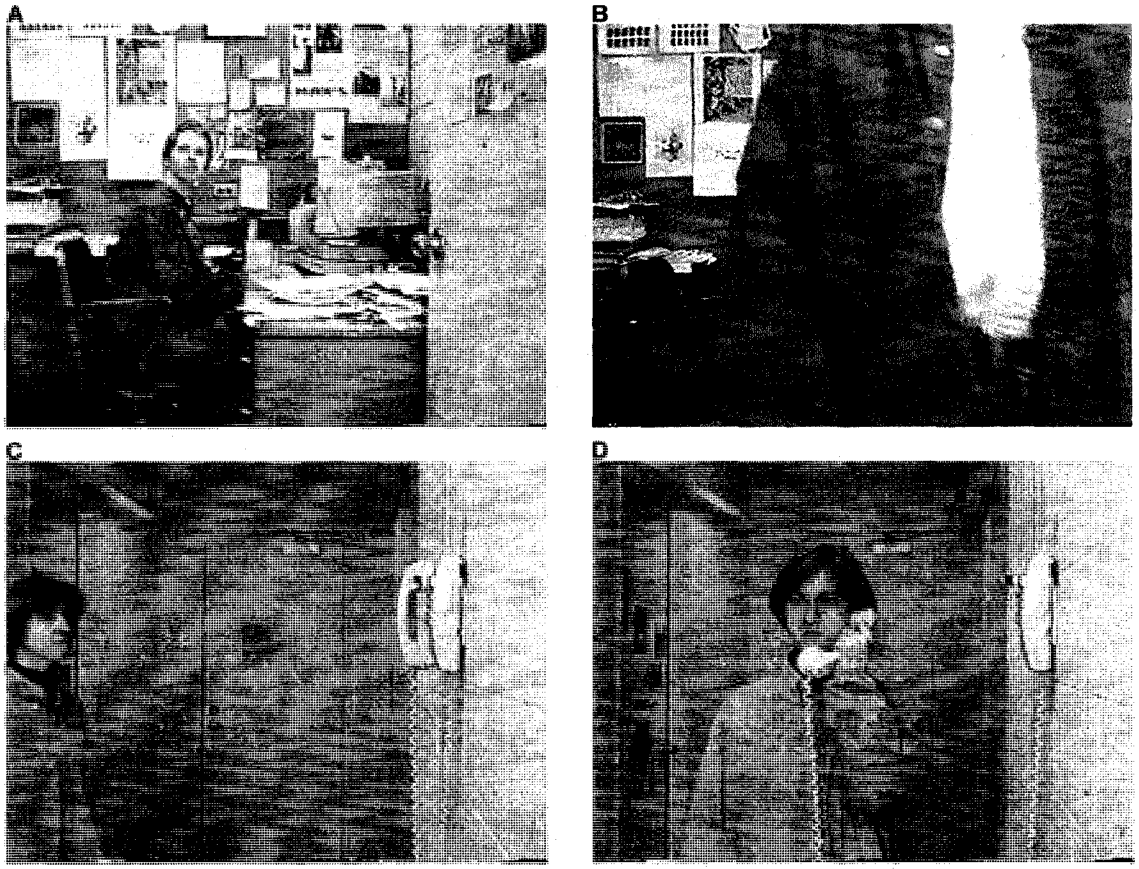


Figure 31. Sample frames from the short movie by Levin and Simons (1997, p. 504). In this movie an actor was sitting at a desk and heard the phone ring (a); got up from the desk and moved towards the door (b); the camera then cut to a view of the hallway, at which point a different actor walked to the telephone (c); and answered it (d).

examined whether unsuspecting pedestrians would notice a change to their conversation partner during a real-world interaction. As illustrated in Figure 32, while the experimenter asked directions to a pedestrian, the interaction was rudely interrupted by two men carrying a door. While hidden by the door, the two experimenters switched places. The original experimenter walked away with the door, while the new experimenter stayed behind and continued the conversation with the pedestrian. Later, when asked whether they had noticed anything unusual, only 47% of the pedestrians reported the change in experimenters. Interestingly, all the pedestrians who had accurately detected the switch were similar in age to the experimenters.

In an attempt to explain this pattern of results, they conducted a similar experiment, with the exception that the experimenters were dressed as construction workers. The obtained verbal reports from this study indicated that only 33% of young students noticed the switch of experimenters. Overall these studies demonstrate that the phenomenon of change blindness is not an artifact produced by artificial disruptions, but can also occur under more naturalistic conditions. Moreover, Simons and Levin (1998) hypothesized that successful change detection appears to have been dependent on social group membership. Pedestrians sharing the same social group as the experimenters tended to expend effort encoding visual features that could later be used to differentiate the experimenters from other members of their social group, or in this case from each other. By contrast, when pedestrians did not share the same social group as the experimenters, they seem to have only formed a representation of the experimenter's social group category (e.g., young student, construction worker), abstract information which was insufficient to distinguish the two versions of the experimenter. Thus, lending

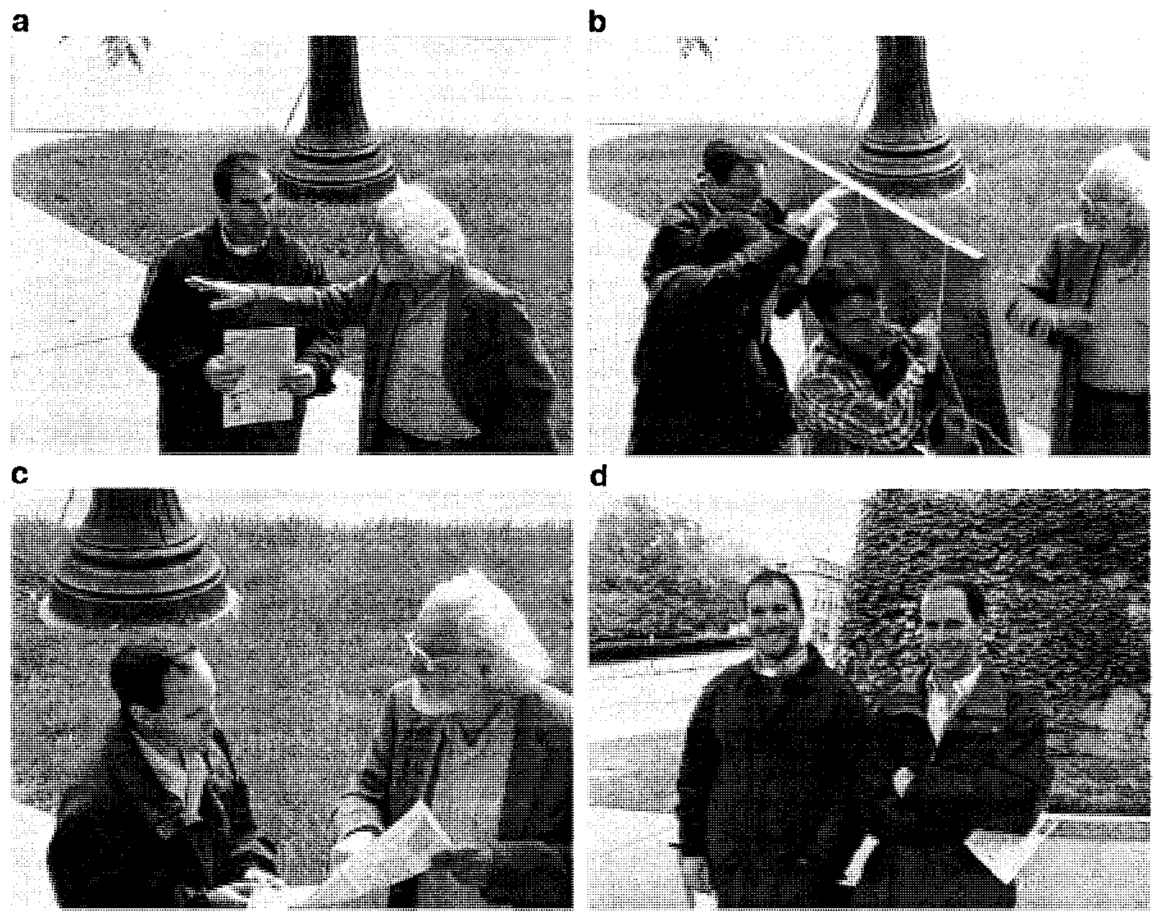


Figure 32. Sequence of events in Simons and Levin's (1998) study (p. 646). Experimenter 1 asked a pedestrian for directions (a); after 10-15 seconds of conversation two other experimenters carrying a door rudely passed between them, during which Experimenter 1 grabbed the back of the door carried by Experimenter 2 (b); Experimenter 2 stayed behind and continued to ask directions to the unsuspecting pedestrian (c).

support to the notion that successful change detection also requires effortful processing of the attended objects.

The assumption in all of the studies reviewed so far was that visual disruptions were necessary to produce change blindness in order to divert attention away from the location of the change signal. Simons, Franconeri and Reimer (2000), however, hypothesized that scene changes could also go unnoticed in the absence of visual disruptions, provided that they were sufficiently gradual to not cause large change signals that capture attention. In a series of studies, they compared change detection performances for object additions/deletions and color changes, with and without the use of visual disruptions. In the gradual condition the original image was dissolved into the changed image over a 12 second period. By contrast, in the disruption condition, the original image and modified image were interspersed by a blank screen. In both conditions the modified image remained visible until the participants indicated the location of the change via a click of the mouse.

The participants' responses indicated no statistically significant difference in the detection of object additions/deletion between the gradual and disruption conditions (64% vs. 57%, respectively). Color changes, however, were found to be detected less often than addition/deletion changes, especially in the gradual change condition (31% vs. 41%). Overall, Simons et al. (2000) concluded that these results demonstrated that change blindness can occur even in the absence of a visual disruption. According to Simons et al. (2000), the pattern of results suggest that in intentional change detection tasks observers implement different strategies depending on whether they are searching for gradual changes or changes across a disruption. In the disruption condition, faced with the knowledge that they will be required to compare the two images, participants may have placed a greater emphasis on encoding and remembering the visual details of the original image. On the other hand, participants in the gradual change condition may have believed that the visual disturbances caused by the scene change would automatically draw attention and, therefore, did not require the formation of a detailed mental representation.

Many possible explanations for the change blindness phenomenon have been proposed. The most popular explanations are described in detail by Simons (2000), which include: (1) *Overwriting*, that is visual information that is not encoded from the original display is replaced in the representation by new information (e.g., Rensink et al., 1997; Rensink, 2000a, b); (2) *First impressions*, that is visual details from the original display are encoded to obtain the scene gist, and provided the scene meaning remains constant across views the details of the new scene will not be encoded (e.g., Friedman, 1979); (3) *Nothing is stored*, that is no or very little information is stored internally, rather the external world acts as a memory store (e.g., O'Regan et al., 1999); (4) *Nothing is compared*, that is detailed representations of the two displays are formed, but not compared unless triggered by inconsistent meaning across views or by the experimenter (e.g., Simons, 1996; Levin & Simons, 1997; Simons & Levin, 1998); and (5) *Feature combination*, that is partial representations from each view are combined to form a new coherent representation of the visual scene (no evidence for this explanation exists in the change blindness literature, but has been found in eyewitness recognition research, such as Loftus (1979)).

In a series of studies Simon, Levin and their colleagues (e.g., Simons, Chabris, Schnur & Levin, 2002; Levin, Simons, Angelone & Chabris, 2002; Angelone, Levin & Simons, 2003; Beck & Levin, 2003; Mitroff, Simons & Levin, 2004; Varakin & Levin, 2006) attempted to determine the cause of change blindness, by examining whether the observers had preserved mental representations of the pre- and post-change stimuli when the change was not detected. By asking participants to complete 4-item line-ups, two-alternative force-choice tasks, and scene descriptions, Simons, Levin and their colleagues were able to measure participants' recognition accuracy rates for the pre- and post-change actors or objects. In real-world situations and motion pictures, Simons, Levin and colleagues provided evidence that change blindness can result from either the absence of a sufficient representation of the scene; a failure to retain the representation after forming it; or a failure to compare the retained representations of the pre- and post-change information. Simons, Levin and colleagues found that although some observers were initially blind to the change, when subsequently asked about the change many were able to recall and provide a rich description of the changed item. By contrast, some studies demonstrated that accurate change detection was associated with higher recognition rates for the pre- and post-change items (i.e., nothing is stored in non-detectors), whereas others found comparable performance between successful change detectors and non-detectors (i.e., nothing is compared in non-detectors). Simons, Levin and colleagues explained these inconsistent findings as being the result of the observer's expectations about the experimental situation, which in turn influenced how the scenes were encoded and represented. For example, awareness that they were taking part in an experiment, or that they would later be questioned about what they saw, may have resulted in participants intentionally encoding many visual details that would normally go unrepresented in real-world situations.

Finally, Simons and Levin's research group proposed that the failure of observers to compare pre- and post-change information occurs when the meaning of the scene remains constant across views or a comparison is not triggered by the experimenter. Empirical evidence for this proposal can be found in a series of studies by Henderson and Hollingworth, who demonstrated that changes to semantically inconsistent target objects (e.g., a fire hydrant in a living room scene) in scenes tended to be detected more accurately than changes to semantically consistent target objects (e.g., a chair in a living room scene) (Hollingworth & Henderson, 2000; Hollingworth, Williams & Henderson, 2001; and Hollingworth & Henderson, 2003). Henderson, Hollingworth and their colleagues proposed that the visual system encodes the details of the scene to derive the scene's meaning (i.e., gist), which is then relied upon across views. Observers may refer back to a few features to assure that they are in the same scene, but with the underlying assumption that objects or people do not drastically change from one moment to the next, efforts are not wasted re-encoding all the details with each view of the scene.

Change Blindness and Eye Movements

Although the change blindness studies reviewed in the previous section have provided us with valuable information, they all suffer from one significant limitation - none provide direct evidence that the participants had been fixating, or had previously fixated, the changed location prior or during the scene change. The reported change blindness phenomenon may simply reflect situations in which the change to the target object occurred in the observer's visual periphery where visual acuity is diminished. Henderson and Hollingworth (1999b) were the first to examine eye movement patterns during saccade-contingent changes. The changes to the target object involved either an instantaneous rotation by 90° around its vertical axis or a disappearance from the scene, and occurred either during the first saccade "toward" or "away" from the target object. A control condition was added in which the change occurred during a saccade to a non-target object. Participants were asked to view the computer-rendered color images of naturalistic scenes for a future memory test and indicate when they detected a change to any of the objects in the current scene.

Detection rates indicated that participants were better at noticing object deletions than object rotations, in particular when the deletion occurred during the saccade "toward" the target object. Henderson and Hollingworth (1999b) proposed that in the "toward" condition visual attention had been allocated to the saccade target location prior to the saccade and, therefore, preferentially encoded, retained or compared this information across saccades. Overall, fixating the target object resulted in higher rotation and deletion detection rates than when the target object was neither fixated immediately before nor after the saccade. Together, these findings suggest that fixating the object immediately before or after it is changed is necessary for change detection, but not sufficient. These findings were later replicated by O'Regan, Deubel, Clark, and Rensink (2000), in a study in which participants failed to notice the change in more than 40% of the trials, even when they were directly fixating the location of the change. Furthermore, Henderson and Hollingworth (1999b) found that successful change detection usually increased as the distance between the location of the fixation immediately before or after the change and the changing object decreased (i.e., visual eccentricity), results which were also replicated by O'Regan et al. (2000). However, in the rotation condition, even when the target object was in near-foveal vision participants were found to be poor change detectors, suggesting that visual acuity alone could not account for change blindness. Analysis of scan paths (as shown in Figure 33) demonstrated that in some cases changes were initially missed, but then subsequently detected when participants re-fixated the location of the change. Henderson and Hollingwoth (1999b) concluded that in some cases change blindness is not the result of poor internal representations of the scene, but rather a failure to consult the information until the changed objects has been re-fixated and re-attended (i.e., a comparison failure).

In a similar study, Hollingworth, Schrock and Henderson (2001), were interested in examining whether fixation position also influenced change detection in the "flicker" paradigm. Contrary to the findings in their previous study, they found no significant difference in detection rates between the two change conditions (i.e., deletion/additions and 90° object rotations). Participants, however, were quicker at detecting deletions/additions than the 90° object rotations, because object deletions and additions tend to create larger visual disruptions than rotations. Eye movement data indicated that at the time of response and final change (prior to the response) participants were more likely to be fixating the region of the target object, as compared to any other region of the visual scene. In addition, participants were quicker to fixate the target object when it was changing than when it was not, especially in the rotation condition. Hollingworth et al., (2001) concluded that together these results suggest that fixation position played a key causal role in successful change detection, in particular for the more difficult changes in the rotation condition.

However, because these findings could also be attributable to the orienting of

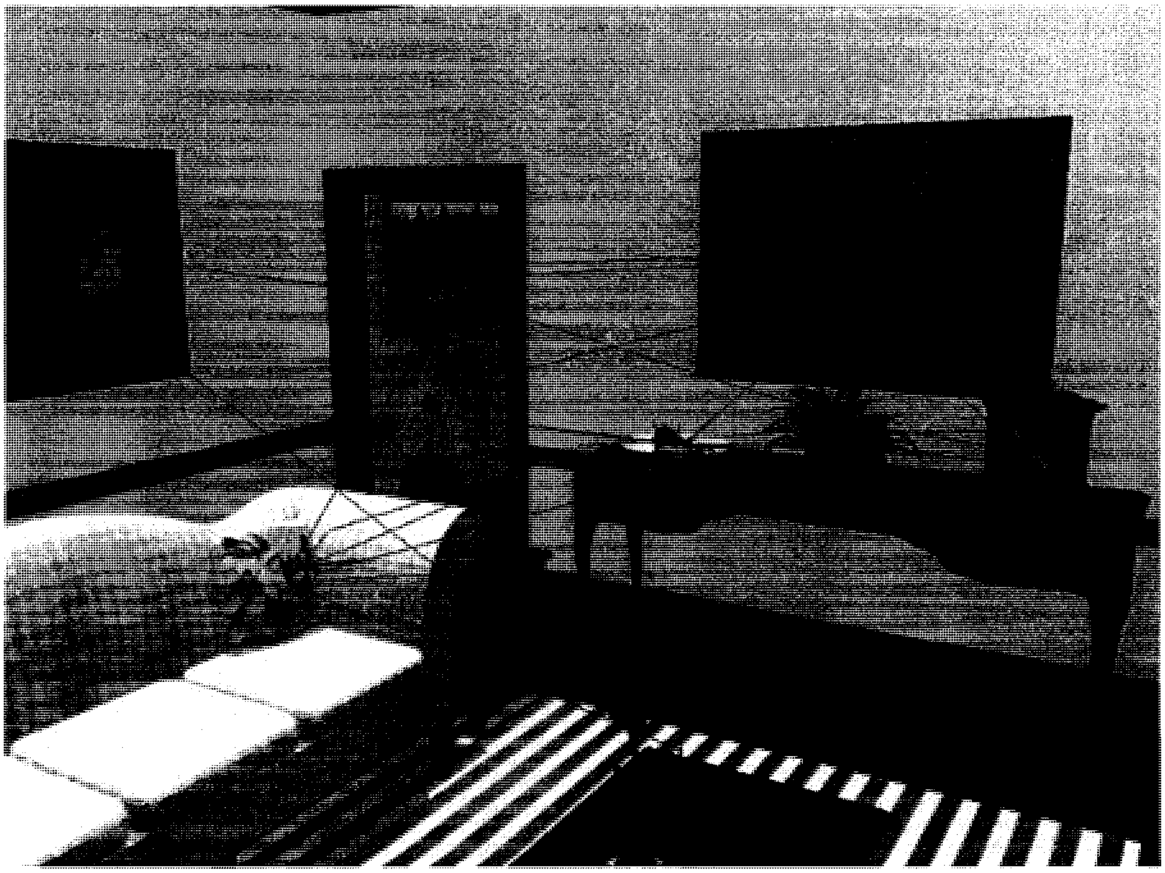


Figure 33. Sample scene and scan pattern of a participant in the "toward" change condition from Henderson and Hollingworth's (1999, p. 439) study. The green lines represent eye movements made before the change, whereas the purple lines represent eye movements made after the change. In this case, the participant did make an eye movement to the clock prior to its change and was blind to the scene change, a 90° rotation of the clock,

visual attention, which is believed to be closely linked with fixation position (e.g., Henderson, 1996), a second experiment was conducted. In this experiment, participants were either free to move their eyes, or were asked to keep their eyes focused on the center of the screen. The results from this study demonstrated that participants were more accurate and quicker at detecting changes, and produced fewer false alarms (i.e., perception of change when none was made), when the target objects could be attended and fixated (i.e., free movement), compared to when they could not be fixated and only attended covertly (i.e., no movement). Hollingworth et al. (2001) concluded that although fixating the changing object appears to be an important factor for change detection, it is not the only factor, as participants were still able to detect changes to the target objects, although more poorly, without moving their eyes.

Analogous to the studies by Simons, Levin and colleagues (e.g., Simons, Chabris, Schnur & Levin, 2002; Levin, Simons, Angelone & Chabris, 2002), Hollingworth and Henderson (2002) examined the nature of scene information retained within long-term memory (LTM). Once again using the saccade-contingent change detection task (i.e., change before or after, or no change), the target objects were either replaced by another object from a different category (i.e., type change; e.g., knife to a fork) or from the same category (i.e., token change; e.g., bread knife to a butcher knife). Following the change detection test participants were administered a forced-choice discrimination test. Hollingworth and Henderson (2002) not only found that change detection performance without prior fixation of the target was no better than false alarm rates, but that performance increased as participants spent more time fixating the target object before the change. Thus, earlier reports of change blindness may have occurred, in part, because of a failure to fixate the target object before the change. Furthermore, participants were better at detecting type changes than token changes, either because of qualitative differences in how information was encoded or the greater visual similarity between objects that share the same category (i.e., token change). In addition, performance on the discrimination task was found to be better than the change detection rates, suggesting that the change detection measure may underestimate the nature of our visual representations in LTM, which is in fact fairly detailed. Finally, similar to the findings in their earlier study (Henderson & Hollingworth, 1999b), analyses of participants' scan paths indicated that the majority of accurate detections resulted from a re-fixation to the location of the target object after it had changed. Hollingworth and Henderson (2002) hypothesized that re-fixation may have cued retrieval of the previously encoded information for subsequent comparison to the current perceptual information.

The possibility that poor change detection occurs as a result of a failure to compare views, was further tested in Hollingworth's 2003 study. As shown in Figure 34, Hollingworth (2003) used a novel approach to divert attention from the location of the change signal. A neon green dot was flashed briefly in the scene, prior to the onset of the patterned mask that was inserted between the original and changed scene. The target object was changed by either rotating it 90° in depth or by swapping it with another member from the same basic level category. In half of the trials a cue was introduced (i.e., an arrow which pointed toward the target object) after the target change had occurred, in order to examine the effects of limiting retrieval and comparison to a specific region of the scene. Overall, participants were found to be more accurate and quicker at detecting changes to the target object in the post-cue condition than the no-cue condition.

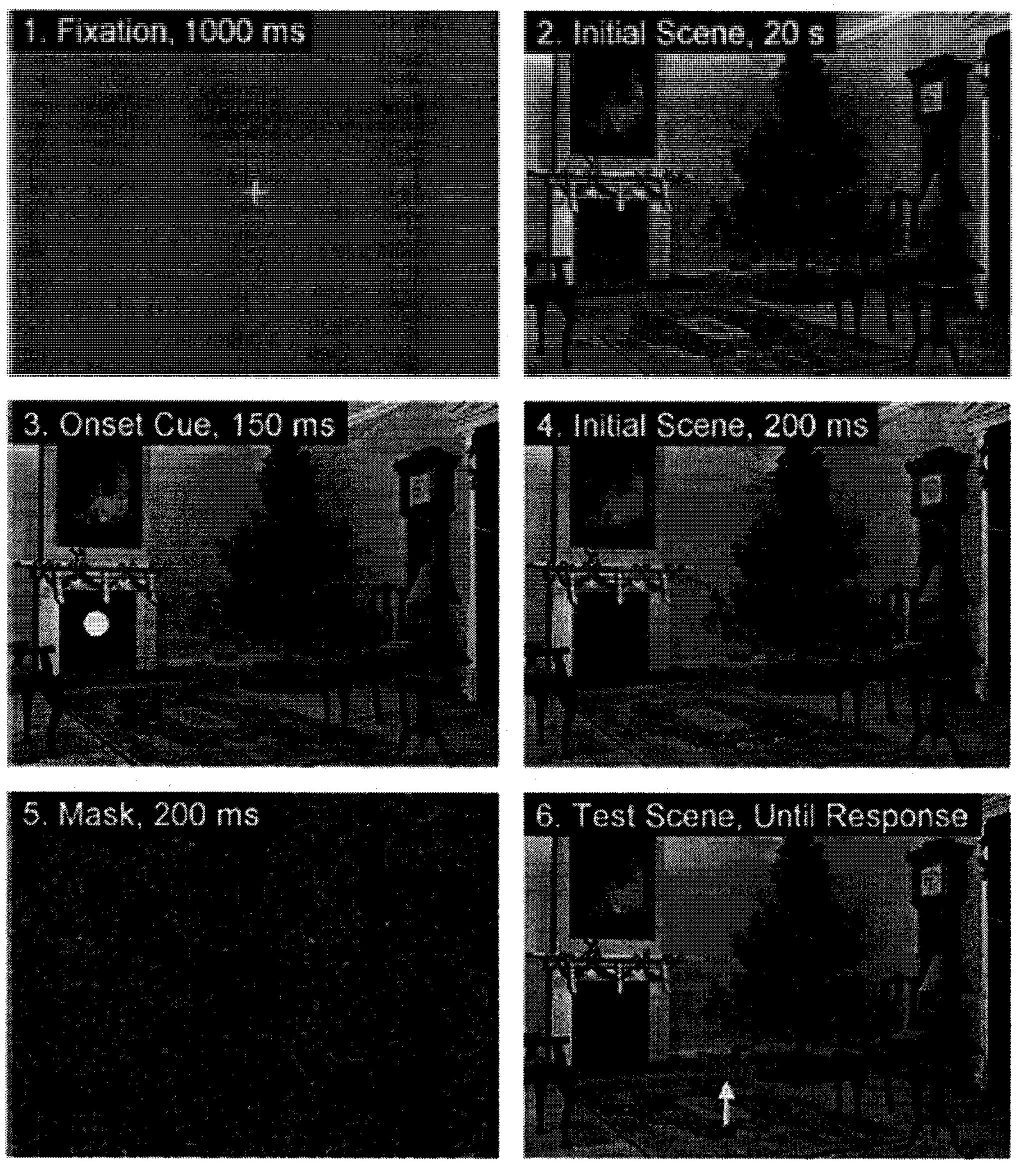


Figure 34. Sequence of events in a trial for Hollingworth's (2003, p. 392) study. In this case the target object was the gift under the tree, which rotated 90° in depth.

These results, Hollingworth (2003) argued, lend support to the notion that change blindness does not likely arise from an inability to retain rich visual information across saccades, but rather from a failure to retrieve and compare the pre- and post-change visual representations. Once again, however, this factor does not fully account for the change blindness phenomenon, because even when the target object was fixated before and after the change, a considerable number of changes (46% and 70% for type and token changes respectively) were undetected (Hollingworth & Henderson, 2002).

Following this series of studies, Hollingworth and Henderson (2002, also reviewed in Henderson & Hollingworth, 2003; Hollingworth, 2003), proposed an alternate theory of scene perception, which was labeled the "visual memory" theory. The theory, which has some similarities with Kahneman, Treisman and Gibbs's (1992) "objects file" theory, suggests that over multiple fixations fairly detailed visual representations of attended objects are formed within LTM. These visual representations do not contain sensory information but rather abstracted visual, conceptual and semantic information. These abstracted representations are "indexed to a position in a map coding the spatial layout of the scene, forming an object file" (Hollingworth & Henderson, 2002). Hollingworth and Henderson (2002) argued that after attention has been withdrawn from the object, visual representations can only be maintained very briefly within vSTM. In conclusion, the retrieval of LTM object files in order to compare this information with the current perceptual representations (i.e., change detection), occurs by re-attending the position in the scene at which information about the object was originally encoded. Hollingworth and Henderson's "visual memory" theory differs from Rensink's "coherence" theory in 3 important ways: (1) changes to previously attended objects can still be detected after attention has been withdrawn, because visual representations have been preserved, rather than rapidly overwritten; (2) fairly detailed visual representations can be retained in LTM, which allows for late detections of earlier unseen changes; and (3) attended objects form more abstract visual representations, than detailed point-by-point representations.

In summary, the findings from the change detection tasks, recognition and discrimination tasks, as well as eye movement scan paths, seem to generally converge onto some significant points. In order to successfully detect scene changes, the target objects need to be previously attended and fixated, followed by a re-fixation to the location of the objects after the change. Direct probing (e.g., direct questioning, recognition tasks, post-cues) and re-fixation of the previously attended object, appear to trigger a comparison between the retained visual representation and the current perceptual information in the scene. Successful change detection tends to vary as a function of the time spent focally attending the target object before the change, which possibly reflects the observer's effortful encoding of the scene. More time attending the target objects may allow for greater conceptual consolidation of the representation into LTM, which in turn is then less susceptible to interference from subsequent conceptual information (Potter, 1976).

Finally, the research findings suggest that orienting focal attention, as well as effortful encoding of visual information, appear to be influenced by several factors, such as: an observer's expectations about the situation, task demands, individual differences, scene complexity, the salience of an object and its semantic consistency within the scene. On the other hand, what can be inferred from change blindness research regarding the nature of our internal representations does not seem to converge as nicely. The overt examination of observers' recall for pre- and post-change stimuli and scan paths appears to indicate that change detection measures underestimate the richness of our mental representations, which may actually be very detailed. Contrary to popular belief, however, these internal representations are unlikely to have a photographic quality, which in most instances is not necessary since viewers usually have the real thing at their disposal.

2. Monitoring Eye Movements during Sentence Comprehension using Dynamic Displays and a Change Detection Task

The main purpose of the second set of studies was to further examine the time-course of integration between linguistic representations and representations computed by the visual and conceptual systems, but with the use of complex dynamic scenes. More specifically, the aim of this study was to investigate how the semantic-selection properties of different classes of verbs guide attention and eye movements toward referents of their grammatical objects and instruments during sentence comprehension, and how these may be further constrained by information provided by the visual context. The goal was also to assess the nature of participants' visual representations, with the use of an implicit change blindness and long term recognition task, and evaluate their influence on eye movement measures. Moreover, the measures obtained from these implicit tasks allowed us to expose the underlying cognitive processes that are not necessarily revealed by the eye movement recordings (i.e., covert attention directed to the target stimulus without movement of the eye). Finally, it allowed for the never reported study of the change blindness phenomenon in complex dynamic scenes, and the use of linguistic post-cues in possibly limiting comparison of the pre- and post-change representations.

This set of studies consisted of one normative study, two pilot studies, and two experiments. In the normative study, data were collected on preferred object and instrument nouns for a set of denominal and non-denominal verbs (Study 1). To select the most effective change detection task and motion pictures (i.e., equally difficult to detect, yet not impossible, scene changes), change detection scores were measured and compared across two change blindness paradigms (Study 2) and multiple movies (Study 3). In Experiment 2a and 2b participants' eye movements were tracked as they watched dynamic scenes in which, unbeknownst to them, an item could disappear. Experiment 2a and 2b were identical, except that in the latter the movies were accompanied by a sentence describing the visual scene. Experiments 2a, thus, provided baseline measures with which to compare the data obtained in Experiment 2b. Performance differences across the two experiments were then used to make inferences regarding the nature of the interaction between the visual and linguistic systems.

2.1 Normative Study: Verb-Complement Selection

A paper-and-pencil task was conducted in order to select the preferred object and instrument nouns for a set of denominal and non-denominal verbs.

2.1.1 Method

2.1.1.1 Participants

Thirty-eight Concordia university undergraduate students participated in this experiment. They were all native speakers of English. All received course credit for their participation.

2.1.1.2 Materials and Design

The paper booklet had a total of 71 fill-in the blank sentences, with 33 restrictive denominal sentence frames (e.g., "The woman will *vacuum* the ______ with the ______") and 38 less-restrictive non-denominal sentence frames (e.g., "The man will *carve* the ______ with the ______") (see Appendix A4).

2.1.1.3 Procedure

After completing the consent form (see Appendix B3), participants were asked to fill in the blanks with objects that best fit the context of the sentence. Participants were further instructed that in some cases the object that best matched the verb could also be the same as the verb in the sentence (e.g., "The man will *telephone* the doctor with the *telephone*"). This study took approximately 20 minutes to complete.

2.1.2 Results and Discussion

Of the initial 71 plausible sentence frames, 16 sentences of each verb type (i.e., denominal and non-denominal) were chosen. The following two criteria were used to select the final set of 32 sentences: (1) the denominal verbs had strongly preferred noun complements (e.g., *carpet* for *vacuum*) and instrument nouns identical to the verb (e.g., …with the *vacuum* for *vacuum*); and the non-denominal verbs had strongly preferred noun complements (e.g., *turkey* for *carve*) and instrumental cases non-identical to the verb (e.g., …with the *knife* for *carve*); (2) the verb's grammatical object and instrument referents could be pictured by concrete basic-level objects, rather than abstract hard-to-picture objects (e.g., *snow*), and could be digitally edited to disappear with ease. When the most frequent objects provided for each verb did not meet these criteria, the second or third most frequent noun was selected (e.g., *canoe* instead of *boat* for *paddle*).

Sentences were then created to include a main clause of the form NP1 (Noun Phrase1)-*will*-Verb-NP2. The NP1 always made reference to the agent in generic form (e.g., *the man*, *the woman*, etc.) and the NP2 made reference to the target object in the scene (e.g., *vegetables, carpet*). The denominal verbs in the sentences made reference to the target instrument in the scene (e.g., *vacuum*), whereas for the non-denominal verbs an instrumental case was added after NP2 (e.g., "*with the knife*"). In order to render the denominal sentences realistic (i.e., as we do not commonly say "The woman will *vacuum* the carpet with the *vacuum*") and of similar length to the non-denominal sentences, a description of the location in which the action took place was added after NP2 (e.g., "The woman will vacuum the carpet *in the basement*"). All sentences were also elongated by adding an initial patch clause, which were always of an adverbial type (e.g., *Later, after doing the laundry…*), followed by the main clause. The purpose of this was twofold: (1) to allow time for the items to disappear before the onset of the verb in the sentences (as described in the main experiments below); and (2) to serve as a script of the agent's present action (e.g., "*doing the laundry*") and future action with the target items (e.g., "*later, after*…the woman will *vacuum*…") in each scenario. To avoid capturing participants' visual attention towards the changing items, the agents within each movie scene performed tasks that did not involve manipulating the target items. Thus, the future actions referred to in each sentence, which involved interacting with the target items in the scene, were never observed by the participants.

2.2 Pilot Studies

2.2.1 Change Blindness Paradigm Selection

As previously mentioned, change blindness has been shown to occur using multiple techniques: (1) saccade-contingent (e.g., Grimes, 1996; Henderson et al., 1999; Hollingworth et al., 2002; Henderson et al., 2003); (2) blink-contingent (O'Regan et al., 2000); (3) blank screens (e.g. Rensink et al., 1997; 2000; Hollingworth et al., 2001); (4) "mudsplashes" (O'Regan et al., 1999); (5) a cut or pan in a motion picture (Levin et al., 1997; Simons, 1996; Angelone et al., 2003); (6) a real-world disruption (e.g. Simons et al., 1998; Levin et al., 2002; Simons et al., 2002); (7) a neon green dot (Hollingworth, 2003); and (8) gradual dissolve (Simons et al., 2000). The latter change detection paradigm seems the most appropriate for the purposes of our study, for several reasons: (1) it would produce the least amount of visual disturbance; (2) it maintains visual continuity; (3) it can be used in combination with the visual world paradigm (e.g., one could not track eye movements in a real-world disruption paradigm, as well as perfectly synchronize the events with the sentences); (4) it would more closely resemble the dynamic visual displays used by de Almeida et al. (in preparation); (5) it may be less disruptive for eye movement recording; and (6) it is most ecologically valid (although items in our environment do not regularly dissolve).

To our knowledge, the "gradual dissolve" paradigm has never been employed in combination with dynamic naturalistic scenes. Although the stimuli used by Simons et al., (2002) were described as movies, in actuality they consisted of a pair of images in which one slowly dissolved into the other (e.g., a house in a field with and without a chimney). Furthermore, the gradual change from one image to the other occurred over a period of 12 seconds, too long of a delay for the purposes of our study. These methodological modifications may produce a strong transient that causes participants' attention to be drawn to the location of the change, thus, rendering the detection task unusable for the purposes of this study. A pilot study was, therefore, conducted to explore this possibility, by comparing the detection of changes occurring during a "flicker paradigm" to the detection of the same changes when they occurred during the modified "dissolve paradigm."

2.2.1.1 Methods

2.2.1.1.1 Participants

Twenty Concordia university students participated in this experiment for course credit or monetary compensation. They all had normal or corrected-to-normal vision.

2.2.1.1.2 Materials and Design

Based on the 32 sentences selected from the normative study described above, 32 unique scenarios were created and filmed (see Appendix A5). The digital movies were produced and edited by two professional film producers. Each movie was 171 frames (i.e., 5.73 seconds) in length and consisted of indoor or outdoor naturalistic scenes. Film resolution was set at 720 x 480 pixels in MPEG-4 video format at 30.0 frames per second. No camera movements or zooms were performed, and the only source of motion within each movie was that of the agent performing a given action. The movies were produced at a large furniture store (Ikea), where the store displays were arranged to resemble common household rooms (e.g., kitchen scenes filled with appliances and cooking utensils). The rest of the scenes were produced in an experimenter's house. The agents were always at the center of the image performing a simple task (e.g., opening the fridge and getting a bottle of cola), with the two target items (i.e., depictions of the verb's grammatical object and instrument) on each side of the agent, at approximately equal distances.

Of the 32 movies, 10 were selected for the purpose of this pilot study, and digitally edited to form the two change versions. In the "dissolve" condition, after 30 frames (i.e., 1 second), either the target object (e.g., candles) or the instrument (e.g., matches) slowly dissolved over a period of 65 frames (i.e., approximately 2.17 seconds) (see Figure 35). In the "flicker" condition, similar to the methodology used by Rensink (Rensink et al., 1997, 2000), a blank grey screen of 3 frames (i.e., 0.1 seconds) was added every 8 frames (i.e., approximately 0.27 seconds) of the movie to form a flicker effect (see Figure 36). Contrary to the typical "flicker" paradigm, however, the target item (object or instrument in the scene) only disappeared once, during the gray screen of the 94-96th frame (i.e., 3.13–3.20 seconds). This assured that the same target item, in both change paradigms, had completely disappeared from the screen by approximately the same time within the trial (i.e., 95th frame).

There were a total of 20 movies (10 movies X 2 versions), which were divided into two groups. Each participant viewed one list containing 10 experimental trials, 5 trials of each change paradigm, and 10 filler trials in which nothing disappeared. Each participant saw only one of the two possible versions of each scenario (i.e., flicker or gradual dissolve). Moreover, for the experimental trials, the disappearing item was counterbalanced between the left and right side of the agent.

2.2.1.1.3 Procedure

The movies were displayed on a Sony Trinitron Multiscan E500 21" monitor (75 Hz refresh rate) placed 57 cm in front of the participant. The movies were presented in a ProShow Gold slideshow. After completing the consent form (see Appendix B4), participants were instructed to watch the 20 movies carefully and after each movie, when

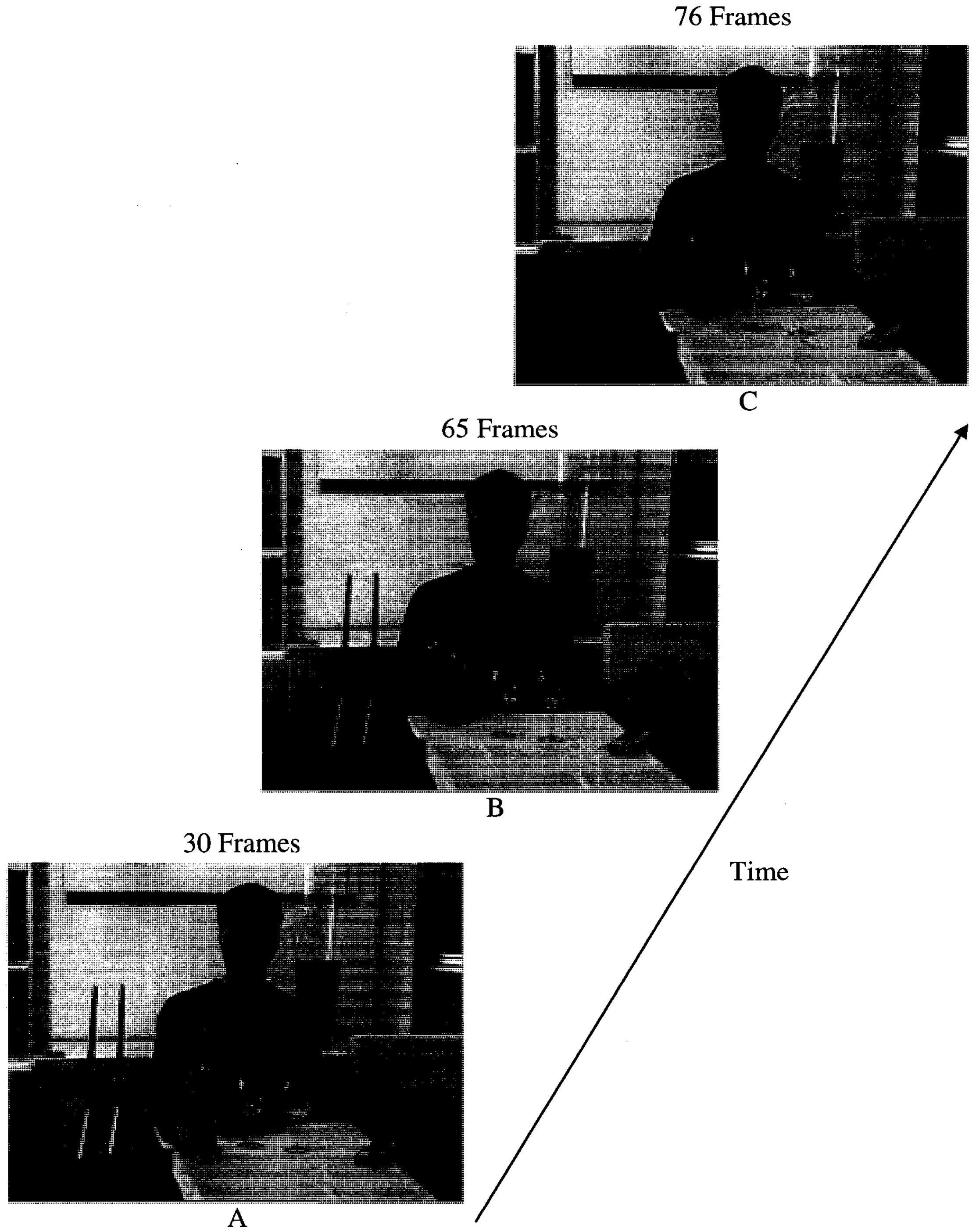


Figure 35. Time sequence of movie events for dissolve condition: (A) movie played for 30 frames unchanged; (B) either the object or its associated instrument slowly dissolved over 65 frames; and (C) movie continued for an additional 76 frames without the presence of the dissolved item.

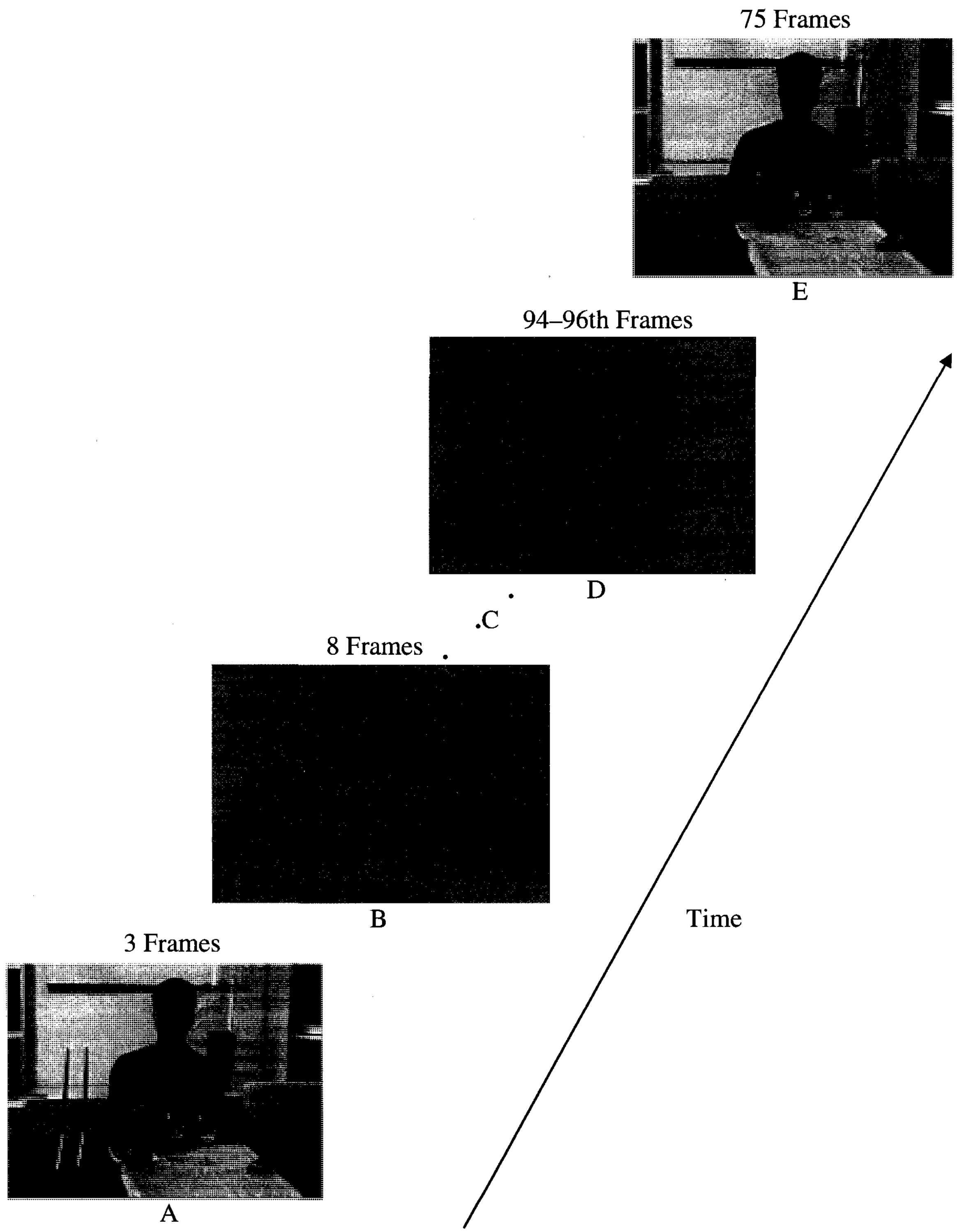


Figure 36. Time sequence of movie events for flicker condition: (A) movie played for 3 frames; (B) a grey screen appeared for 8 frames; (C) A and B alternated repeatedly; (D) during the 94-96th grey frame the object or its associated instrument was removed; and (E) the movie continued for an additional 75 frames without the presence of the removed item.

the screen turned black, describe in as much detail as possible what they had just seen in the movie. For each movie, participants had 20 seconds to write down their description of the scene in a paper booklet provided by the experimenter. The participants were not informed that in 50% of the trials an item on the screen would disappear. This study took approximately 10 minutes to complete.

2.2.1.2 Results and Discussion

Overall, participants detected approximately the same proportion of scene changes in the "gradual dissolve" (*mean* = 30%, *SD* = 14.14%) condition and the "flicker" condition (*mean* = 27%, *SD* = 12.61%). The change detection rates were analyzed using two *t*-tests, a paired samples t-test for the by subject [*t1*] analysis and an independent samples *t*-test for the by item [*t2*] analysis. There was no statistically significant difference between the two change conditions, $t1(19) = .946$, $p = .658$; $t2(18) = .449$, $p = .356$. Since the modified "gradual dissolve" paradigm produced similar change detection rates as the "flicker' paradigm, for the reasons mentioned above, this methodology was used for the main experiments in the second set of studies.

2.2.2 Movie Selection

The entire set of 32 movies was then digitally edited using the same methodology as described above (see Figure 35). In order to ensure that item disappearances in each scenario were of similar saliency, a pilot study was conducted in which participants were asked to rate the noticeability of each scene change. This was of particular importance, since more salient scene changes might quickly alert participants to the underlying purpose of the study, particularly if, by chance alone, these scenes happened to be displayed early within the experiment. Once aware of the possibility of scene changes, participants may become more attentive to scene details and, therefore, more efficient change detectors. Consequently, overall change detection rates would vary greatly across participants as a function of how early they received the more salient trials.

2.2.2.1 Methods

2.2.2.1.1 Participants

Thirty Concordia university students participated in this experiment for monetary compensation or course credit. They all had normal or corrected-to-normal vision.

2.2.2.1.2 Materials and Design

The 32 movies were digitally edited to form three change versions (a detailed description is provided above in the normative study section): after the initial segment of 50 frames (i.e., 1.67 seconds), either the verb's direct object (e.g., candles, in the future accompanying sentence "…the man will light the *candle* with the match") or its associated instrument (e.g., "matches") appeared to dissolve gradually over a period of 50 frames (i.e., 1.67 seconds), whereas in the control condition, the movies were left intact. Thus, there were a total of 96 movies (32 scenes X 3 versions) and each were 180 frames (i.e., 6 seconds) in length.

2.2.2.1.3 Procedure

The movies were displayed on a Sony Trinitron Multiscan E500 21" monitor placed 57 cm in front of the participant. The movies were presented in a slideshow using the ProShow Gold program. After completing the consent form (Appendix B5) participants were instructed to carefully watch a series of movies in which an item may disappear. Participants viewed all 3 versions of each movie, and each trial lasted 240 frames (i.e., 8 seconds). As illustrated in Figure 37, each trial began with a red cross

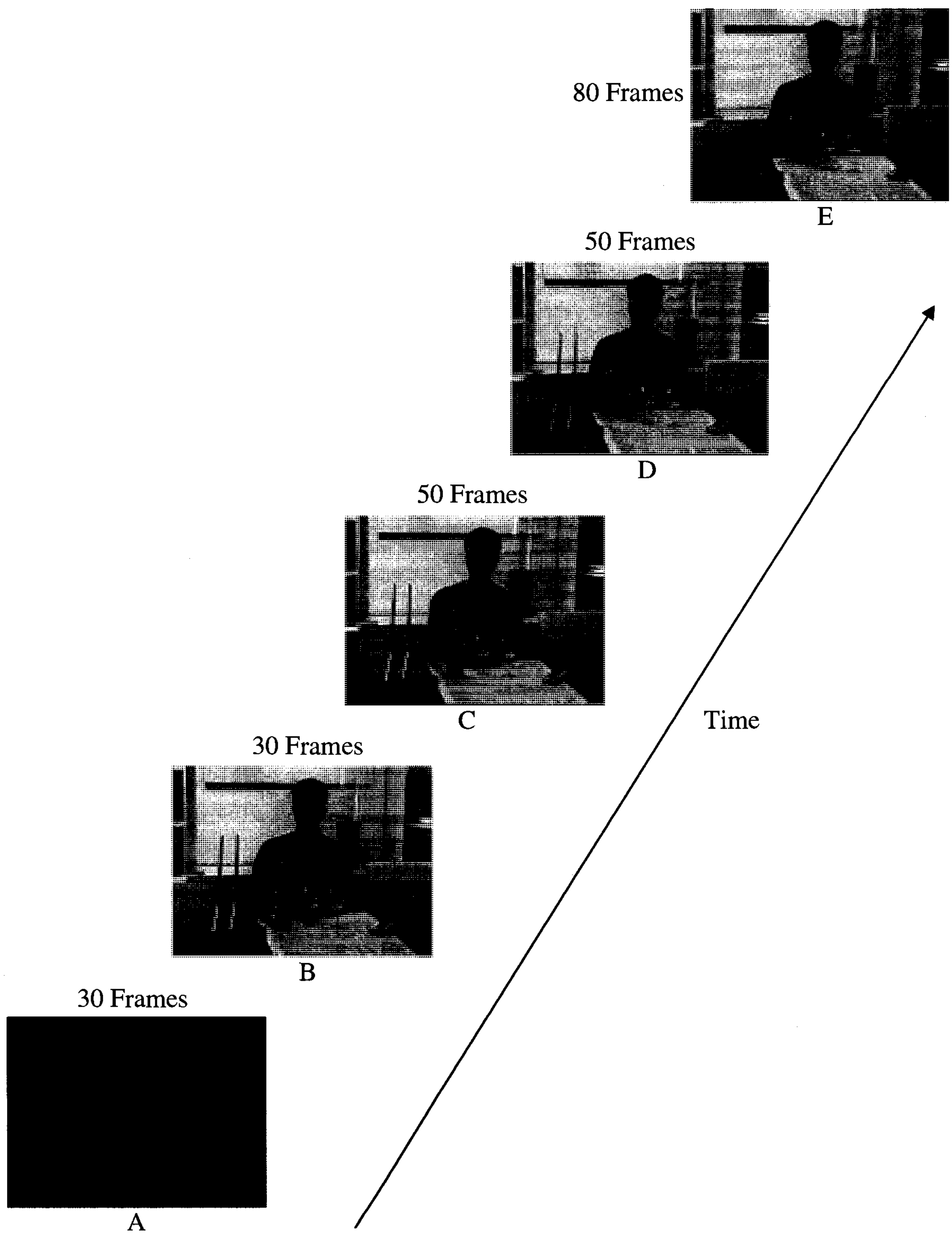


Figure 37. Time sequence of movie events for Experiment 2a: (A) Fixation cross was shown for 30 frames; (B) still of first movie frame was shown for 30 frames; (C) movie played unmodified for 50 frames; (D) object or its associated instrument slowly dissolved over 50 frames; and (E) movie continued without the dissolved item for an additional 80 frames.

displayed for 1 second at the center of a black background. The black background was then replaced by the first still frame of the movie. After 1 second, the cross disappeared (at which point participants could now move their eyes), and the movie began. After 6 seconds the movie stopped, a black screen appeared and participants had 3 seconds to rate in a paper booklet, from a scale of 1 to 5, how noticeable the disappearance was – with "1" indicating that the disappearance was very noticeable and "5" indicating that the disappearance was not noticeable (see Appendix A6). This study took approximately 20 minutes to complete.

2.2.2.2 Results and Discussion

Of the initial set of 32 movie scenes, 18 unique scenes were selected for the two main experiments (9 denominal and 9 non-denominal verb scenes). 12 of the remaining 14 movie scenes were used as fillers in the main experiments. The following three criteria were used to select the final set of 18 movies: (1) on average both items in the scene (i.e., direct object and its associated instrument) received a low rating for disappearance noticeability; (2) target objects and instruments were counterbalanced between the left and right hand side of the agent; and (3) there was a good variety of items, actors and scene categories (see Appendix A7).

2.3. Experiment 2a: Change Detection Task without Verbal Postcue

2.3.1. Methods

2.3.1.1 Participants

Twenty-seven Concordia university undergraduate students participated in this experiment for course credit or monetary compensation. They were all native speakers of English and had normal or corrected-to-normal vision.

2.3.1.2 Materials and Design

The 18 unique scenes selected above were used for the two main studies (see Appendix A7). As described above, each movie had been digitally edited to form three change versions (object or instrument gradually disappeared, or nothing disappeared (control condition)). Thus, there were a total of 54 movies (18 scenes X 3 versions) that were evenly distributed into 3 lists of materials. Each participant viewed one list containing 18 experimental trials (6 of each change version) and 12 filler trials (where nothing disappeared), for a total of 30 unique movies. These 12 fillers were added in order to prevent participants from anticipating when an item in the scene would disappear. For the recognition and naming task that would follow (described below), both target items were cropped (i.e., 18 objects and 18 instruments) from the first still frame of each movie.

2.3.1.3 Procedure

The movies were presented on a Sony Trinitron Multiscan E500 21" monitor placed 57 cm in front of the participant. The visual stimuli were presented via PsyScope software on an Apple Macintosh G4 computer. Participants' eye movements were recorded using the EyeLink-I head-mounted eye-tracker at a sampling rate of 250 Hz. Although viewing was binocular, for analysis simplicity, only eye movements from the left eye were recorded. Head movements were minimized with the use of a chinrest. Minor head movements were corrected by a system of four LED sensors affixed to each corner of the monitor, and a head-mounted camera that continuously measured participants' head position in relation to the screen.

After completing the consent form (see Appendix B6), participants were instructed to watch the 30 movies carefully. Following each movie, participants were given 15 seconds to verbally describe, in as much detail as possible, what they had just seen in the movie. The participants were unaware at the beginning of the experiment that in 40% (i.e. 12/30) of the trials an item on the screen would gradually dissolve. The sequence of movie events was the same as described in the pilot study above (see Figure 37). The experimenter tape-recorded the participants' responses, which were then transcribed verbatim. Each trial lasted 8 seconds and the full viewing session lasted about 25 minutes, which included eye-tracker calibration time. Finally, without prior warning, the participants were then asked to view a Microsoft PowerPoint slideshow, in which 72 cropped images were shown one by one (see Figure 38). The slides consisted of the 18 target objects and instruments in each scene (i.e., 18 scenes x 2 targets = 36), as well as 36 filler images that did not appear in any of the previously viewed movies. Participants were then asked to indicate in a paper booklet, by a "yes" or "no", whether they recalled seeing the currently displayed item in any of the previously viewed movies (recognition task), as well as identify it (identification task) (see Appendix A8).

Because the two experiments had similar methodologies, the materials, design and procedures used for the first experiment (Experiment 2a) are described in detail, followed by the modifications made to Experiments 2b. The results section, which comprises the analyses conducted on both experiments is collapsed into one comprehensive section and presented after the two methodology sections. This type of layout was chosen to allow for greater ease in comparing performance across the two experiments.

2.4. Experiment 2b: Change Detection Task with Verbal Postcue

The same materials and procedures were used in this experiment with the

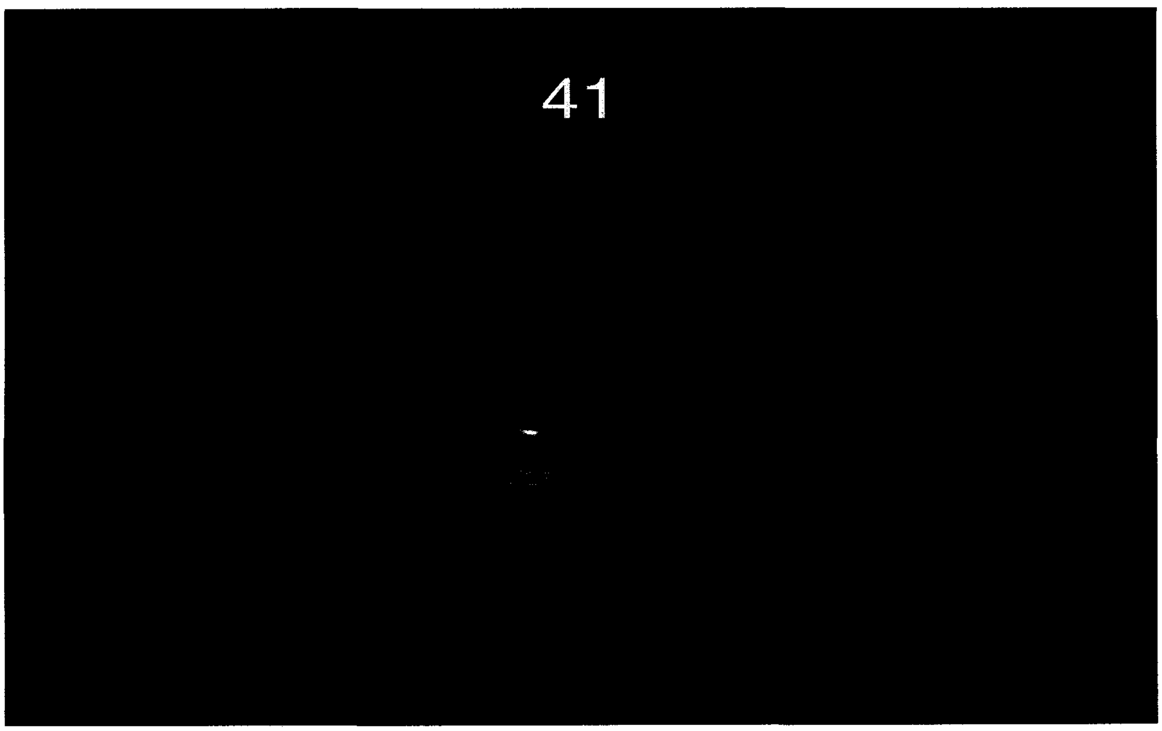


Figure 38. Sample PowerPoint slide for the post-test recognition task.

exception of an aurally presented sentence which described the visual scene.

2.4.1. Methods

2.4.1.1 Participants

Twenty-seven Concordia university students participated in this experiment for course credit or monetary compensation. They were all native speakers of English and had normal or corrected-to-normal vision.

2.4.1.2 Materials and Design

The sentences (see Appendix A7 & A9) were recorded by a female student at a natural pace, and as described in the normative study above, consisted of two different types of sentence contexts, one with a restrictive denominal verb (e.g., "...the man will *whisk* the eggs... "), and another with a less-restrictive non-denominal verb (e.g., "...the woman will *toss* the salad with the tongs").

2.4.1.3 Procedure

As illustrated in Figure 39, the procedure was the same as the one adopted in Experiment 2a, with the exception that the sentences and films were synchronized such that verb onsets in the sentences corresponded to complete disappearance of target items in the scene (i.e., 160^{th} frame or 5.33 seconds).

2.5 Results

Several sets of analyses were conducted on participants' verbal and written responses, as described in the following paragraphs.

Proportion Correct Change Detection of Target Items. The proportions of correctly detected changes were calculated for the three change conditions (i.e., no-change control, object or instrument disappears) across the two verb classes (i.e.,

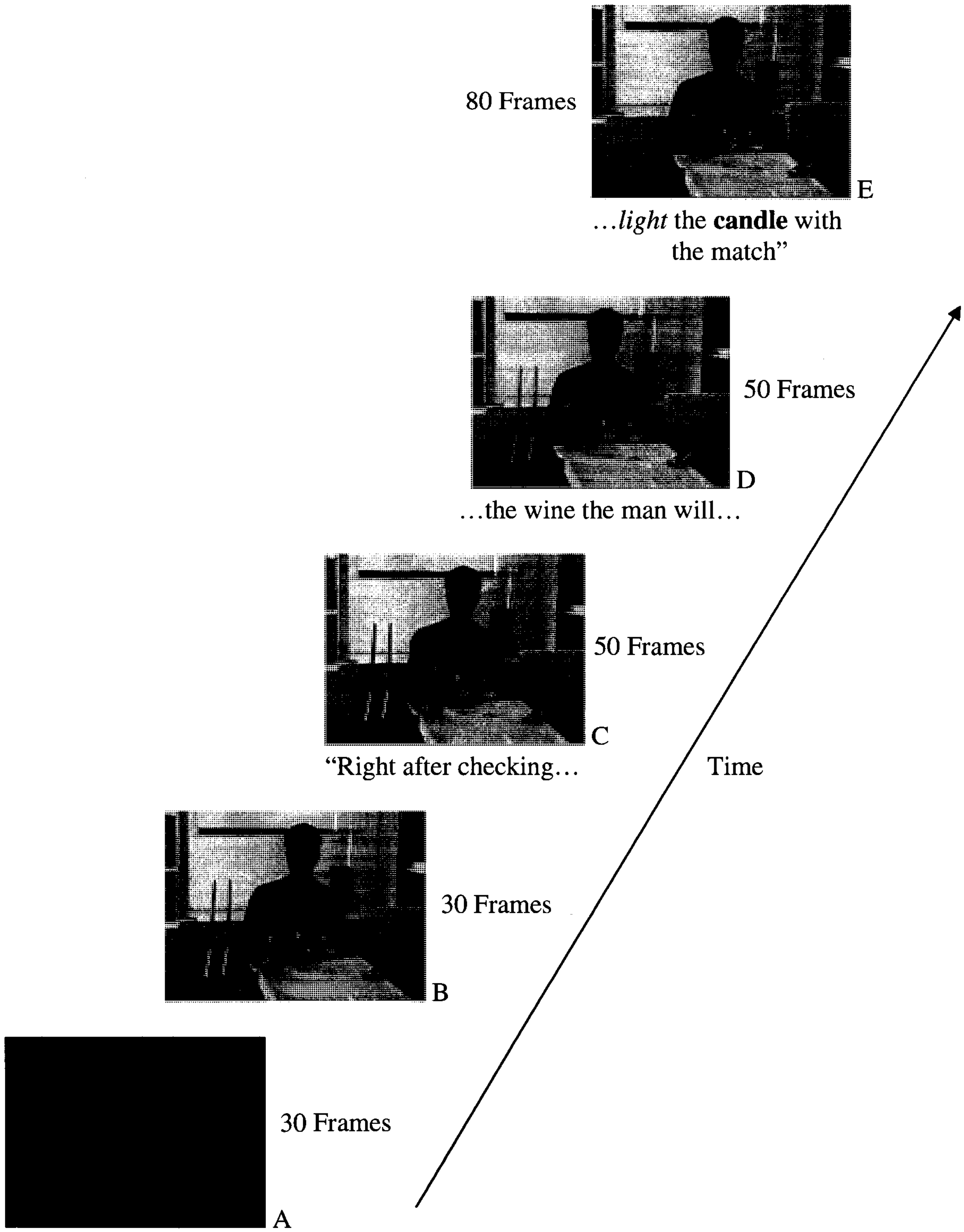


Figure 39. Time sequence of movie events for Experiment 2b: (A) Fixation cross was shown for 30 frames; (B) still of first movie frame was shown for 30 frames; C) movie played unmodified for 50 frames and sentence began to unfold; D) the object or its associated instrument slowly dissolved over 50 frames; and E) at verb onset the movie continued without the dissolved item for an additional 80 frames.

denominal and non-denominal). This measure was examined to determine the nature of participants' visual representations of the scenes and whether different types of linguistic information would improve change detection. The prediction was that participants would be quite poor at detecting scene changes, but that linguistic information would guide eye movements towards the location of the changed item, and thus increase change detection rates. In addition, the expectation was that once participants became aware of the possibility of scene changes, detection rates would increase.

Proportion of Target Items Mentioned. The proportions of target items mentioned by participants in their verbal responses were calculated for the four categories of items (i.e., denominal object and instrument, non-denominal object and instrument) in the no-change control condition. This measure was studied to determine the visual saliency of the different target items in each scene, as well as verify that participants were attending to the accompanying sentences. The expectation was that more central interest target items, that is, items more likely to be spontaneously reported by participants, would attract more attention and, thus, result in greater change detection rates. Furthermore, it was hypothesized that once aware of the possibility of scene changes participants would become more attentive to scene details and, thus, more likely to mention the target items.

Proportion Correct Identification of Target Items. The proportions of correctly identified target items were calculated for the four categories of items in the no-change control condition. This measure was examined to determine whether undetected scene changes, were in part, the result of participants' inability to identify the target items in the scenes. Moreover, this measure was used to verify that the names assumed by the experimenter to be the main labels for each item were correct. The hypothesis was that poorly identifiable target items would result in lower change detection rates, but that the information contained within the related sentences would diminish these effects, by providing participants with a label for these ambiguous items.

Proportion Correct Recognition of Target Items. The proportions of correctly recognized target items were calculated for the four categories of items, across both the no-change control and experimental (i.e., change conditions) trials. These measures were examined to determine the nature of participants' visual representations of the scenes and their role in the change blindness phenomenon. The prediction was that undetected scene changes were, in part, the result of participants failing to encode the target items in the scenes before they changed. Moreover, the expectation was that the descriptive sentences would improve recognition scores by providing labels to ambiguous items, which then would be more easily consolidated into LTM. Finally, it was hypothesized that once aware of the possibility of scene changes participants would become more attentive to scene details and, thus, more likely to encode the target items, which they would then be more apt at recalling in the post-test recognition task.

2.5.1 Proportion Correct Change Detection of Target Items

Participants' verbal descriptions were examined and coded for mention of a dissolving or disappearing item from the scene. Failure to label changed items (e.g., reported seeing "*something* disappear on the left hand side") or misidentification of changed items (e.g., reported seeing "an *apple* dissolve", when it was a *tomato*), were also considered instances of correct change detection. Mean correct detection rates for the experimental and no-change control conditions; the number of experimental trials to first change detection; and correct change detection at and after the first change detection are displayed in Table 7.

For Experiment 2a and 2b, respectively, mean correct change detection rates were 0.51 and 0.46, and participants required on average 3.6 and 3.57 change trials before correctly detecting the gradual disappearance of the target items in the visual scene. The distribution of correct detection rates ranged from 0 to 0.83 for Experiment 2a and 0 to 0.92 for Experiment 2b. One participant in Experiment 2a and 4 in Experiment 2b failed to detect any of the scene changes and this was later confirmed when the participants were asked directly whether they had noticed any of the items disappear from the screen. In the no-change control condition there were no false alarms, meaning that participants did not report noticing a change when none occurred. Once participants correctly detected a scene change, correct detection performance significantly increased in the subsequent trials, for both Experiment 2a (from 0.39 proportion correct detection to 0.61), $t(26) = -3.16$, $p = .004$, and Experiment 2b (from 0.37 to 0.55), $t(26) = -2.35$, $p = .027$. Thus, once aware of the possibility of scene changes, participants appear to have become more attentive and vigilant to the possibility of subsequent scene changes.

Table 8 summarizes the results of the analyses conducted on the correct change detection rates for the two experiments, and Figure 40 illustrates correct change detection rates across verb types (i.e., denominal and non-denominal) and target item disappearance (i.e., object or instrument). These verb classifications are somewhat arbitrary for the items in Experiment 2a because the dynamic scenes were not accompanied by sentences, as in Experiment 2b; however, this layout facilitated comparison of the data across the two studies. The proportions of correctly detected changes were analyzed using two repeated measures ANOVA. For experiment 2a, one

Table 7. Correct change detection measures, averaged across participants.

	Detection	Number of Trials	No-Change Control	At First Detection	After First Detection
	No Sentence (Experiment 2a)				
Mean	0.51	3.60	1.00	0.39	0.61
SD	0.23	2.29	0	0.27	0.27
	Sentence (Experiment 2b)				
Mean	0.46	3.57	1.00	0.37	0.55
SD	0.009	0.32	0	0.32	0.34

Table 8. Summary of analyses conducted on correct change detection rates, across the two experiments (without and with accompanying sentence).

Study	Measure	Factor	Analysis	p	η^2
No Sentence	Correct Change Detection	Category	$F1(3,78) = 12.19$	.001 *	.319
			$F2(3,24) = 3.65$	.027 *	.313
Sentence	Correct Change Detection	Verb (V)	$F1(1,26) = 7.63$	.010 *	.227
			$F2(1,16) = 2.41$	.140	.131
		Changed Item (I)	$F1(1,26) = 1.43$	.243	.052
			$F2(1,16) = 1.70$	.210	.096
		V x I	$F1(1,26) = 3.31$	.081	.113
			$F2(1,16) = 2.41$	.140	.096
No Sentence vs. Sentence	Correct Change Detection	Verbal Cue	$F1(1,52) = 0.721$	.400	.014
			$F2(1,32) = 1.93$	.175	.057

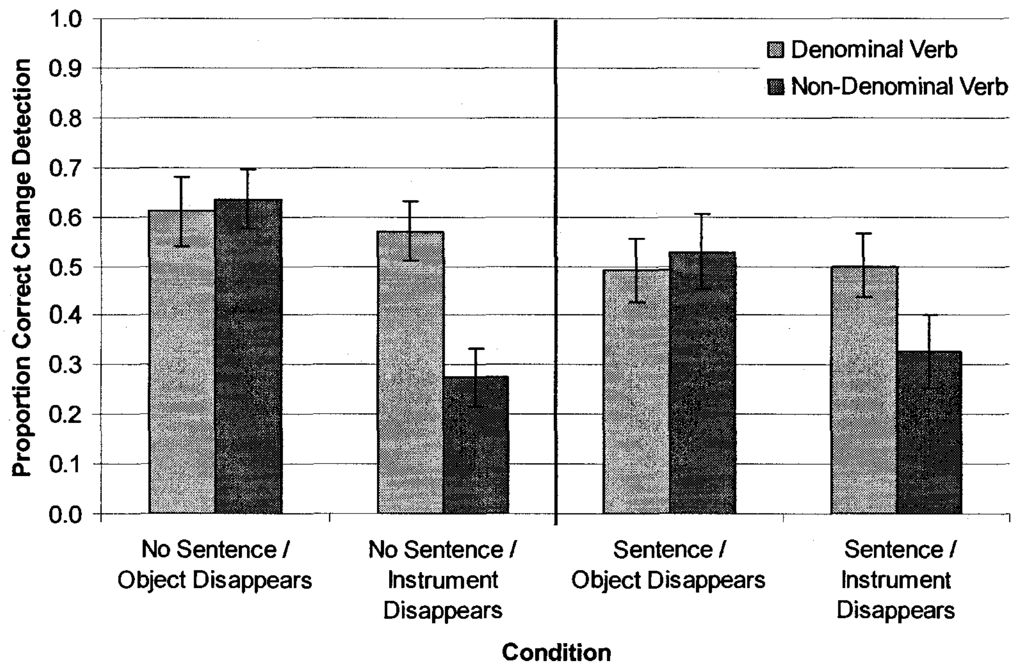


Figure 40. Proportions of correctly detected changes across conditions, without and with the associated sentence.

analysis treated participants as a random effect and category (i.e., denominal object and instrument, and non-denominal object and instrument) as within-subject factors [*F1*], whereas the other treated items as a random effect and category as a within-item factor [*F2*].

As shown in the first panel of Table 8, the analysis revealed a statistically significant main effect of category, and Bonferroni post-hoc pairwise comparisons indicated that the difference between items lay between the non-denominal instrument category (*mean* = 0.27) and the remaining three categories in the by subject analysis (*means* = 0.61, 0.64, 0.57 for denominal objects and instruments, and non-denominal objects respectively), and the non-denominal object and non-denominal instrument categories in the by item analysis. Overall, participants were quite poor at detecting large changes to dynamic scenes in the "gradual dissolve" paradigm, in particular for the items that will later encompass the non-denominal instrument category for Experiment 2b.

For Experiment 2b, the proportions of correctly detected changes were analyzed using one repeated measures and one mixed design ANOVA. One analysis treated participants as a random effect, and verb types and changed target items as within-subject factors [*F1*], whereas the other treated items as a random effect, changed items as a within-item factor and verb type as a between-item factor [*F2*]. As indicated in the second panel of Table 8, the analysis revealed a statistically significant main effect of verb type, for the by subjects analysis only, no main effect of changed item, and no significant interaction between these two factors. Thus, participants' detection rates were lower for the non-denominal target items, and although not significant, this was particularly true for the non-denominal instrument items.

The proportion correct change detection data were then compared across the two studies using one repeated measures and one mixed design ANOVA. One analysis treated participants as a random effect, verb types, changed item, and sentence (presence or absence) as within-subject factors [*F1*], whereas the other treated items as a random effect, changed item as a within-item factor, verb types and sentences as a between-item factors [*F2*]. As shown in the third panel of Table 8, the analysis revealed no significant main effect of sentential context, and no significant interaction between verb and sentence $F1(1,52)= 3.34$, $p= .073$, $\eta_p^2= .060$; $F2(1,32)= 1.01$, $p= .323$, $\eta_p^2= .031$ and between changed item and sentence $F1(1,52)= 0.822$, $p= .369$, $\eta_p^2= .016$; $F2(1,32)= 0.122$, $p= .730$, $\eta_p^2= .004$. Thus, contrary to expectations the different types of linguistic information contained in the accompanying sentences did not significantly decrease the occurrence of change blindness. Overall, the pattern of results was similar across the two studies, with poorer change detection rates for the non-denominal instrument target items. Since the lower detection rates for the non-denominal instrument target items in Experiment 2b are nearly identical to those obtained in Experiment 2a, the findings are unlikely to be the result of any verb effects. Rather, other factors appear to be involved in creating this pattern of results and will be explored in greater detail below.

2.4.2.2 Proportion of Target Items Mentioned in Participants' Verbal Responses

In the no-change control condition, degrees of interest for the target items were assessed by coding participants' verbal responses for any reference to the target objects and instruments, regardless of item misidentification. Table 9 summarizes the results of the analyses conducted on the proportions of target items mentioned in the verbal responses for the two experiments, and Figure 41 illustrates these proportions across verb

Table 9. Summary of analyses conducted on proportions of target items mentioned in participants verbal responses, across the two experiments (without and with accompanying sentence).

Study	Measure	Factor	Analysis	p	η^2
No Sentence	Proportion	Category	$F1(3,78)= 4.94$	.003 *	.160
	Mentioned		$F2(3,24)= 1.37$	.275	.146
Sentence	Proportion	Verb (V)	$F1(1,25)= 2.43$	.132	.089
	Mentioned		$F2(1,16)= 0.88$	.361	.052
		Target Item (I)	$F1(1,25)= 0.261$	.614	.010
			$F2(1,16)= 0.035$	.853	.002
		V x I	$F1(1,25)= 20.3$	.001 *	.447
			$F2(1,16)= 10.2$	.006 *	.390
No Sentence vs. Sentence	Proportion Mentioned	Verbal Cue	$F1(1,51)= 7.04$	.011 *	.121
			$F2(1,32)= 12.5$	.001 *	.280
No Sentence	Proportion	Aware of	$F1(1,7)= 4.67$	.068	.400
	Mentioned	Change	$F2(1,8)= 9.58$	.015 *	.545
Sentence	Proportion	Aware of	$F1(1,6)= 1.41$	.280	.190
	Mentioned	Change	$F2(1,16)= 1.63$	.220	.092

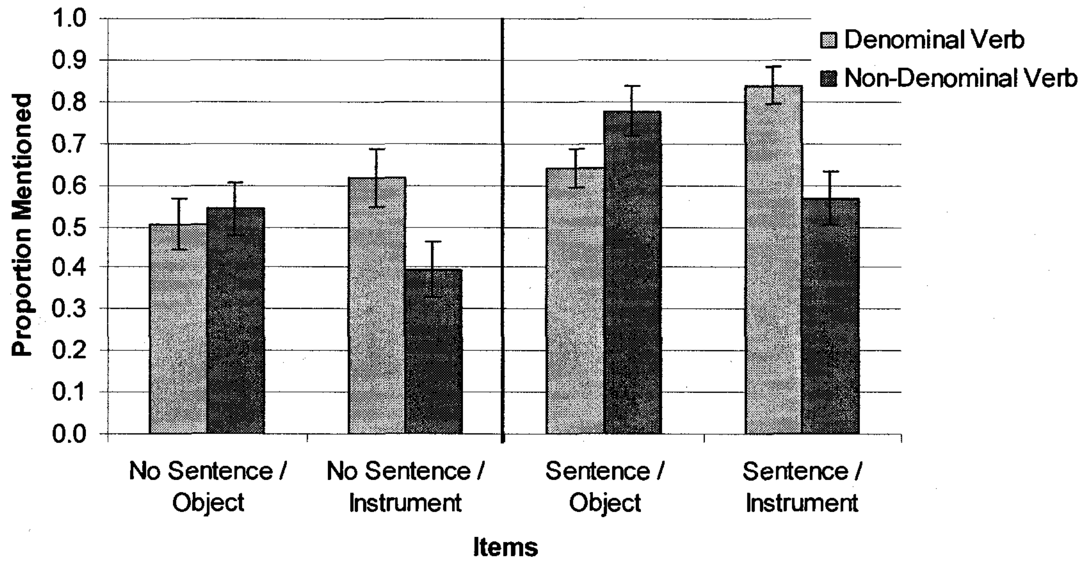


Figure 41. Proportions of target items mentioned by participants in the no-change control condition, across conditions, without and with the associated sentence.

types.

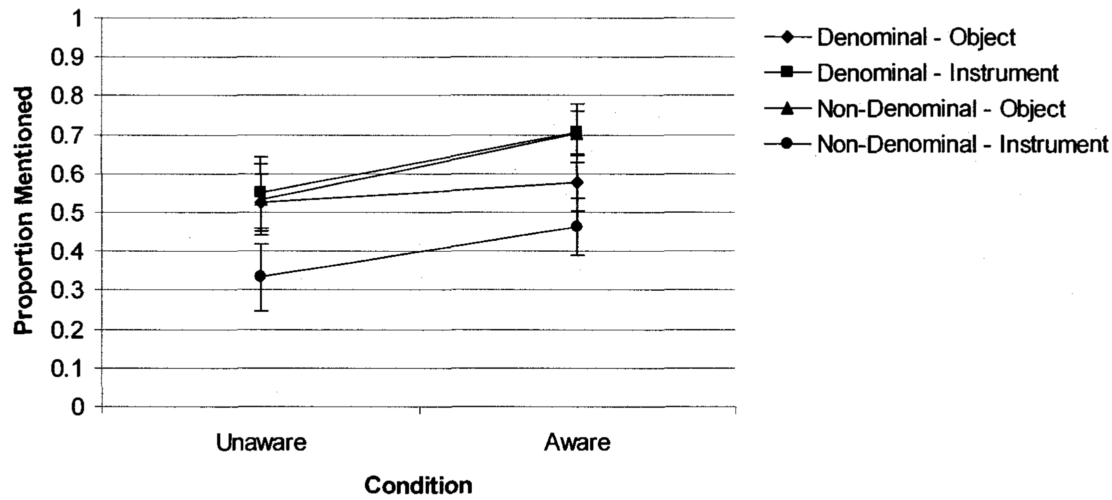
As shown in the first panel of Table 9, for Experiment 2a the two repeated-measures ANOVAs revealed a statistically significant main effect of category for the by subjects analysis only, and the Bonferroni post-hoc pairwise comparisons indicated that the difference between items lay between the denominal and non-denominal instrument items. In general, the target items were mentioned more than 50% of the time, suggesting that they were considered by participants to be of central interest, except for the non-denominal instrument items (e.g., *knife*, in "…the woman will chop the vegetables with the knife") which appear to have attracted less attention. Pearson correlations between the detection rates of the items in each scenario and their corresponding interest grading did not, however, show any statistically significant associations, $r_{DO} = 0.561$, $p = 0.116$; $r_{DI} = 0.257$, $p = 0.504$; $r_{NDO} = 0.124$, $p = 0.752$; and $r_{NDI} = 0.605$, $p = 0.085$. Thus, although participants tended to report the non-denominal instrument items with less frequency in the no-change control trials, it does not appear to have been linked with the lower change detection rates obtained for this category of target items.

As shown in the second panel of Table 9, for Experiment 2b the repeated measures and mixed design ANOVAs revealed no statistically significant main effect of verb type or target items, yet there was a statistically significant interaction between these two factors. Pearson correlations between the detection rates of the target items in each scenario and their corresponding interest grading, however, did not show any statistically significant associations, $r_{DO} = 0.320$, $p = 0.402$; $r_{DI} = -0.217$, $p = 0.576$; $r_{NDO} = -0.125$, $p = 0.748$; and $r_{NDI} = -0.355$, $p = 0.349$. Thus, similar to Experiment 2a participants in the no-change control condition tended to report the non-denominal instrument items with less frequency, but this did not appear to be linked with the lower change detection rates obtained in the experimental trials for this category of target items.

The proportions of target items mentioned by participants in their verbal responses were then compared across the two studies and, as shown in the third panel of Table 9, the analysis revealed a statistically significant main effect of sentential context, but no significant interaction between verb and sentence $\underline{F1}(1,51)= 0.052$, $\underline{p}= .820$, $\underline{\eta_p}^2 = .001$; $\underline{F2}(1,32)= 0.034$, $\underline{p}= .854$, $\underline{\eta_p}^2 = .001$, and between target item and sentence $\underline{F1}(1,51)= 0.009$, $\underline{p}= .923$, $\underline{\eta_p}^2 = .000$; $\underline{F2}(1,32)= 0.005$, $\underline{p}= .946$, $\underline{\eta_p}^2 = .000$. Hence, for Experiment 2b participants were more likely to mention the target items than participants in Experiment 2a, indicating that participants had indeed processed the accompanying sentence. Given that change detection scores were not higher for Experiment 2b, however, these findings do not appear to reflect an increase in target item saliency. Rather, analysis of the verbal responses suggests that participants utilized the accompanying sentence to describe the visual scene. In addition, because the proportions for the non-denominal instrument target items in Experiment 2b were nearly identical to those obtained in Experiment 2a, the lower proportions were unlikely to be the result of any verb effects.

Proportion of target items mentioned by participants in the no-change control condition, before and after first correct change detection, were plotted in Figure 42. As indicated in the bottom panel of Table 9, the two repeated-measures ANOVAs revealed a statistically significant main effect of change awareness, for the by item analysis only, and no significant interaction between category and awareness, $\underline{F1}(3,21)= 1.52$, $\underline{p}= .238$, $\underline{\eta_p}^2 = .179$; $\underline{F2}(3,24)= 0.307$, $\underline{p}= .820$, $\underline{\eta_p}^2 = .037$. A similar pattern was found for

A)



B)

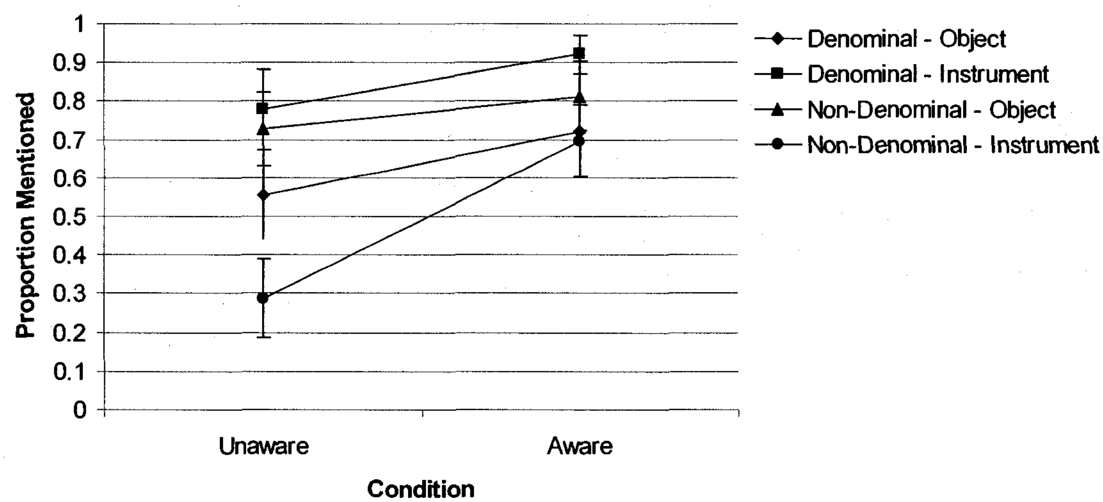


Figure 42. Proportion of target items mentioned by participants in the no-change control condition, before (unaware) and after (aware) first correct change detection, for (A) Experiment 2a and (B) Experiment 2b.

Experiment 2b, but the analysis revealed no statistically significant main effect of awareness, a significant interaction between target items and awareness, for the by items analysis only, $\underline{F1}(1,6)= 0.774$, $\underline{p}= .413$, $\eta_p^2 = .114$; $\underline{F2}(1,16)= 13.24$, $\underline{p}= .002$, $\eta_p^2 = .453$, and no significant interaction between awareness and verb type, $\underline{F1}(1,6)= 1.41$, $\underline{p}= .280$, $\eta_p^2 = .190$; $\underline{F2}(1,16)= 1.37$, $\underline{p}= .259$, $\eta_p^2 = .079$. Overall, the trend in the data indicated that participants' viewing behavior changed when they were no longer naïve to the possibility of item disappearance, as demonstrated by more reporting of the target items in participants' verbal descriptions of the scenes. These findings, thus, support the notion that after the first correct change detection the target items became areas of central interest, which appears to have been particularly beneficial for the originally less attention grabbing non-denominal instrument items.

2.4.2.3 Proportion Correct Identification of Target Items

An inability to recognize the non-denominal instrument items, may have resulted in their poor encoding or retention, and consequently their poor detection when changed. Table 10 summarizes the results of the analyses conducted on the proportions of correctly identified target items for the two experiments, and Figure 43 illustrates these proportions for the object and instrument target items, collapsed over the 3 change conditions and verb types. As shown in the top panel of Table 10, for Experiment 2a the two repeated-measures ANOVA revealed a statistically significant main effect of category for the by subject analysis only, and Bonferroni post-hoc pairwise comparisons indicated that the difference between items lay between the non-denominal instrument category and the remaining three categories in the by subject analysis. Thus, the items within the non-denominal instrument category were found to be less identifiable, as compared to the

Table 10. Summary of analyses conducted on correct identification rates, across the two experiments (without and with accompanying sentence).

Study	Measure	Factor	Analysis	p	η^2
No Sentence	Correct	Category	F1(3,78)= 8.28	.001 *	.242
	Identification		F2(3,24)= 0.471	.705	.056
Sentence	Correct	Verb (V)	F1(1,26)= 1.12	.300	.041
	Identification		F2(1,16)= 0.177	.679	.056
		Target Item (I)	F1(1,26)= 1.04	.317	.039
			F2(1,16)= 0.10	.755	.056
		V x I	F1(1,26)= 44.8	.001 *	.633
			F2(1,16)= 3.70	.073	.056
No Sentence	Correct	Verbal Cue	F1(1,52)= 6.90	.011 *	.117
vs. Sentence	Identification		F2(1,32)= 0.808	.375	.025

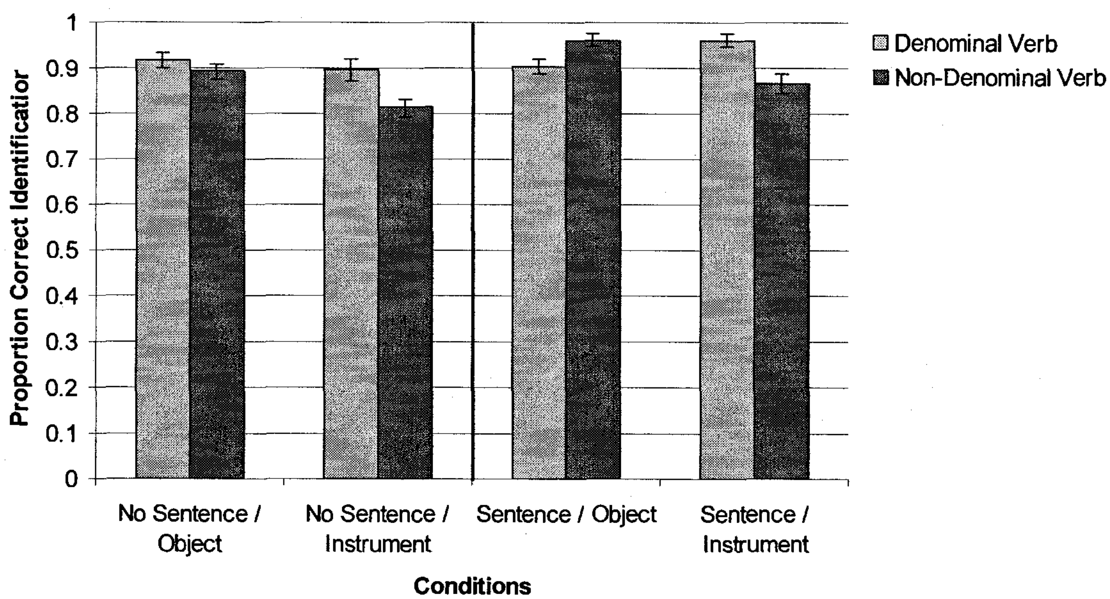


Figure 43. Proportions of correctly identified target items across change conditions, for Experiment 2a and 2b.

items in the other 3 categories.

Similar results were found in Experiment 2b, as shown in the second panel of Table 10, for which the analysis revealed no statistically significant main effect of verb type or target items, yet a significant interaction between these two factors for the by subject analysis only. Thus, for the most part target items were correctly identified by participants (88% and 92% for Experiment 2a and 2b, respectively), however, in both studies the non-denominal instrument items were slightly less identifiable. These results are in line with the findings reported in the previous sections, suggesting that target items within this category that were less identifiable, were also less likely to be mentioned by participants and, therefore, their changes more likely to go unnoticed.

The proportions of correctly identified target items were then compared across the two studies and, as shown in the last panel of Table 10, the analysis revealed a statistically significant main effect of sentential context, no significant interaction between verb and sentence $F1(1,52) = 1.89$, $p = .175$, $\eta_p^2 = .035$; $F2(1,32) = 0.090$, $p = .766$, $\eta_p^2 = .003$, or between target item and sentence $F1(1,52) = 2.14$, $p = .15$, $\eta_p^2 = .039$; $F2(1,32) = 0.362$, $p = .552$, $\eta_p^2 = .011$. Thus, the accompanying sentences, which identified the target items in each scene, appear to have aided participants in correctly labeling each item in the subsequent identification task, and possibly during the experiment. This greater ease at identifying the target items, however, does not seem to have resulted in overall greater scene change detection.

2.4.2.4 Surface Area of Target Items

The surface area (measured in cm^2) taken up by each target item on the screen may explain the lower change detection rates for the non-denominal instrument items.

These values were obtained using the Sante DICOM Editor software, which can calculate the surface area of a specific region in a picture, given a particular screen resolution. Smaller target items may have been less likely to attract participants' visual attention before or during the change (i.e., by causing a smaller change signal), than the larger target items. Figure 44 illustrates the average area sizes for the object and instrument items across verb types. The surface areas were analyzed using a one-way ANOVA and revealed no statistically significant main effect of item size, $F(3,32)= 1.78$, $p= .171$, $\eta_p^2 = .143$.

Surface area, however, may not be the best measure of visual saliency as it does not share a simple linear relationship with apparent size. For instance, an object that takes up double the surface area is not necessarily perceived as being twice as large (Stevens, 1957). Moreover, two objects can have the same area size (e.g., 20 cm^2), but significantly differ in their overall shape; a more compact object (e.g., 4x5 cm) may be more visually salient than a long thin object (e.g., 2x10 cm). In fact when examining the instrument target items that belonged to the non-denominal category, they tended to fall under this latter category (i.e., knife, fork, spoon, ice pick, scissors, stapler, tongs, dish rag, and a box of matches). Thus, the thin elongated target items in the non-denominal instrument category may have been more difficult to identify and less likely to be attended, encoded and retained within short and long-term memory – ultimately, also resulting in lower change detection rates.

2.4.2.5 Proportion Correct Recognition of Target Items

Table 11 summarizes the results of the analyses conducted on the proportions of correctly recognized target items across the two experiments and Figure 45 illustrates

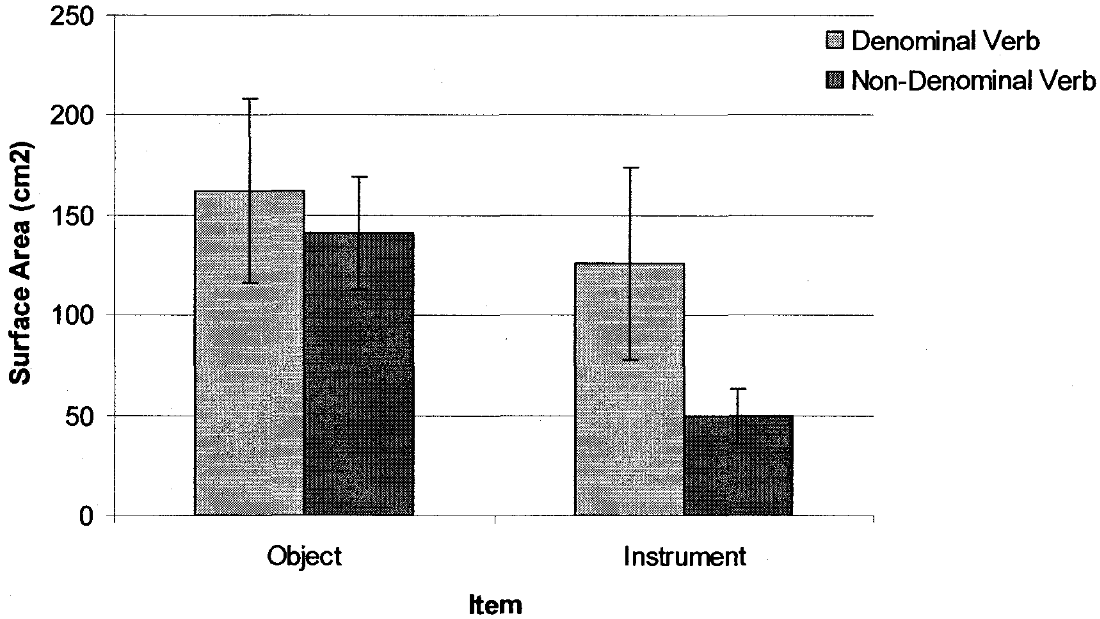


Figure 44. Average surface areas (cm^2) for target object and instrument items across conditions.

Table 11. Summary of analyses conducted on correct recognition rates the for no-change control trials, across the two experiments (without and with accompanying sentence).

Study	Measure	Factor	Analysis	p	η^2
No Sentence	Correct	Category	$F1(3,78)= 2.08$	.110	.074
	Recognition		$F2(3,24)= 1.30$	.297	.140
Sentence	Correct	Verb (V)	$F1(1,26)= 3.14$	.088	.108
	Recognition		$F2(1,16)= 2.26$	.153	.124
		Target Item (I)	$F1(1,26)= 0.00$	.996	.000
			$F2(1,16)= 0.00$	1.00	.000
		V x I	$F1(1,26)= 1.01$	.324	.037
			$F2(1,16)= 0.842$	.372	.050
No Sentence	Correct	Verbal Cue	$F1(1,52)= 2.94$	.092	.053
vs. Sentence	Recognition		$F2(1,32)= 3.26$	.080	.092

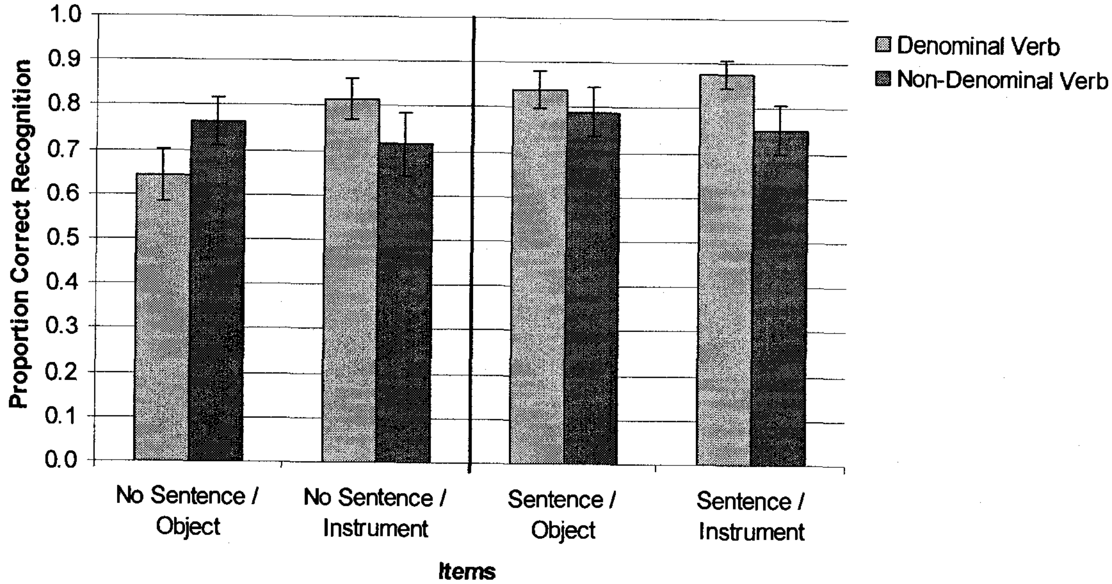


Figure 45. Proportion correct recognition of target items across verb types for the no-change control condition, for Experiment 2a and 2b.

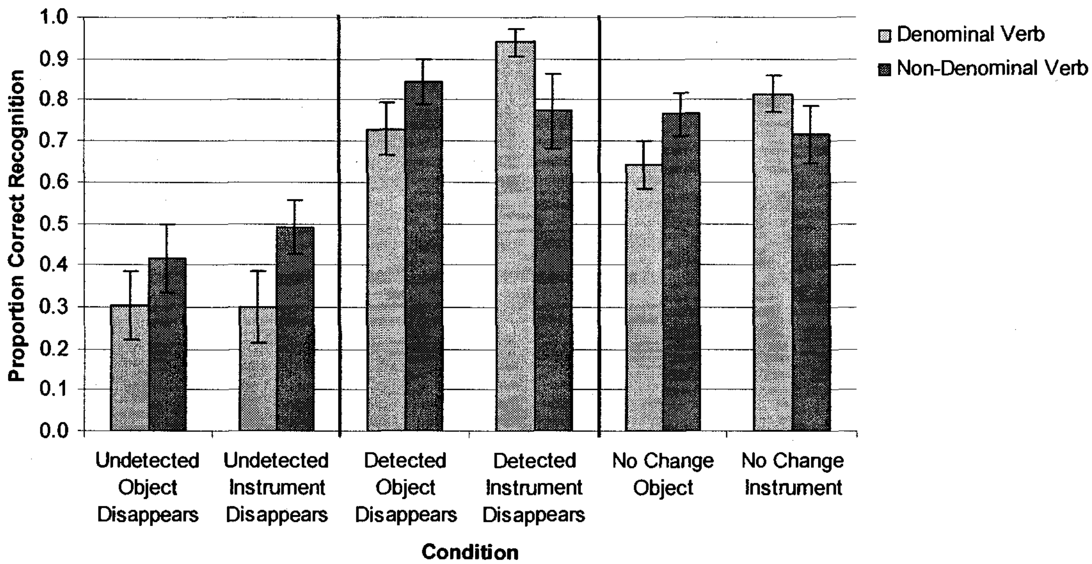
these proportions across verb types for the no-change control condition. As indicated in the top panel of Table 11, for Experiment 2a the two repeated-measures analyses revealed no statistically significant main effect of category. Overall, recognition performances were 63.23% and 69.13% for Experiment 2a and 2b respectively, values significantly above the chance level of 50% correct (*t*(26)= -1495.10, p< .001; *t*(26)= -1760.12, p< .001, respectively), and false alarm rates were under 8% in both experiments (5.35% in Experiment 2a and 7.51% in Experiment 2b). These results suggest that participants, for the most part, had encoded the target items, and were able to retain this information for a relatively long period of time (5-20 minutes later).

As shown in the second panel of Table 11, similar results were found for Experiment 2b, for which the analysis revealed no statistically significant main effect of verb types or target items, and no significant interaction between these two factors. Thus, participants appear to have encoded the majority of the target items in each scene, approximately 73% and 81% for Experiment 2a and 2b, respectively. Furthermore, even though the non-denominal instrument items were less likely to be spontaneously mentioned by participants, and later less accurately identified, they still appear to have been attended, encoded and retained within the visual representations. In comparing proportions of correctly recognized target items from the two studies, as shown in the bottom panel of Table 11, there was no statistically significant main effect of sentential context, or interaction between sentence and verb type, $F1(1,52) = 1.87$, $p = .178$, $\eta_p^2 = .035$; $F2(1,32) = 1.20$, $p = .282$, $\eta_p^2 = .036$, and between sentence and target items $F1(1,52) = 1.00$, $p = .321$, $\eta_p^2 = .019$; $F2(1,32) = 0.594$, $p = .447$, $\eta_p^2 = .018$. Thus, contrary to expectations there was no significant improvement in subsequent recognition scores for the no-change control trials when participants received a verbal description of the scene.

Yet the question remains, what caused the poor change detection rates? The proportion correct recognition scores in Figure 46 for participants who had successfully detected the scene changes, as compared to those who were blind to the change, suggest that one possible reason may have been a failure to encode the target items before they disappeared. Table 12 summarizes the results of the analyses conducted on the proportion correct recognition scores of the target items for the experimental trials across the two experiments. As shown in the top panel of Table 12, for Experiment 2a the two repeated measures ANOVAs revealed a statistically significant main effect of change detection, no significant main effect of target items and a significant interaction between these two factors. Bonferroni post-hoc pairwise comparisons indicated that the difference in correct recognition scores lay between the undetected and correctly detected changed items, and between the undetected and the no-change control items. Hence, participants who had successfully detected the changes to the target items, were later more accurate at recognizing the target items in the post-test recognition task. Or, in other words, participants who had attended, encoded and retained the target items before they disappeared were subsequently better at detecting the change. Whether this encoding occurred prior or while the target item was gradually changing will be revealed by examining the eye movement data.

As shown in the second panel of Table 12, similar results were found in Experiment 2b, for which the analysis revealed a statistically significant main effect of change detection, no significant main effect of verb type and no significant interaction

A)



B)

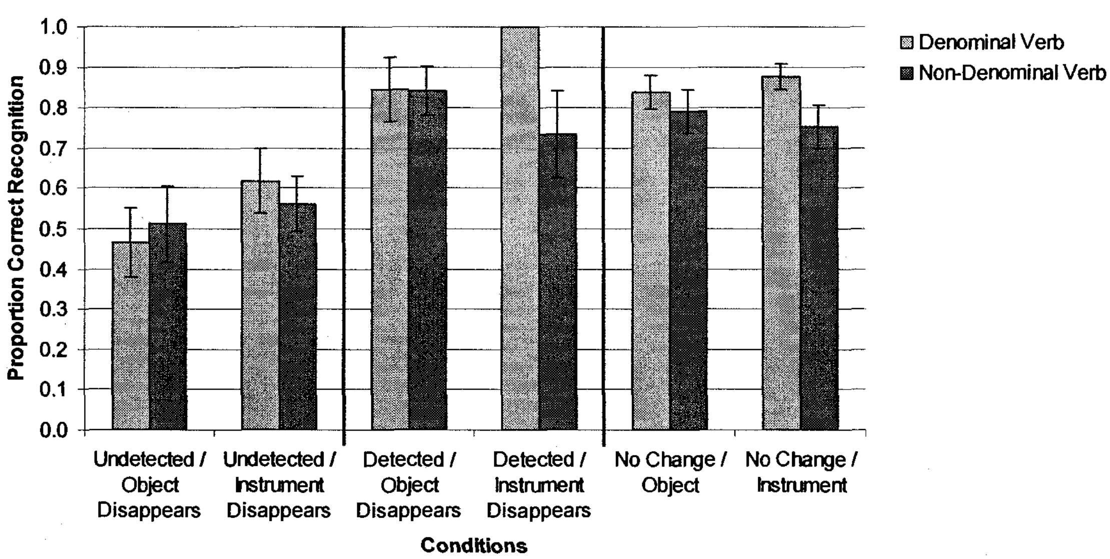


Figure 46. Proportion correct recognition rates for target items when no changes occurred (control condition), and changes were undetected or detected by participants, across conditions, for A) Experiment 2a and B) Experiment 2b.

Table 12. Summary of analyses conducted on correct recognition rates for the experimental trials, across the two experiments (without and with accompanying sentence).

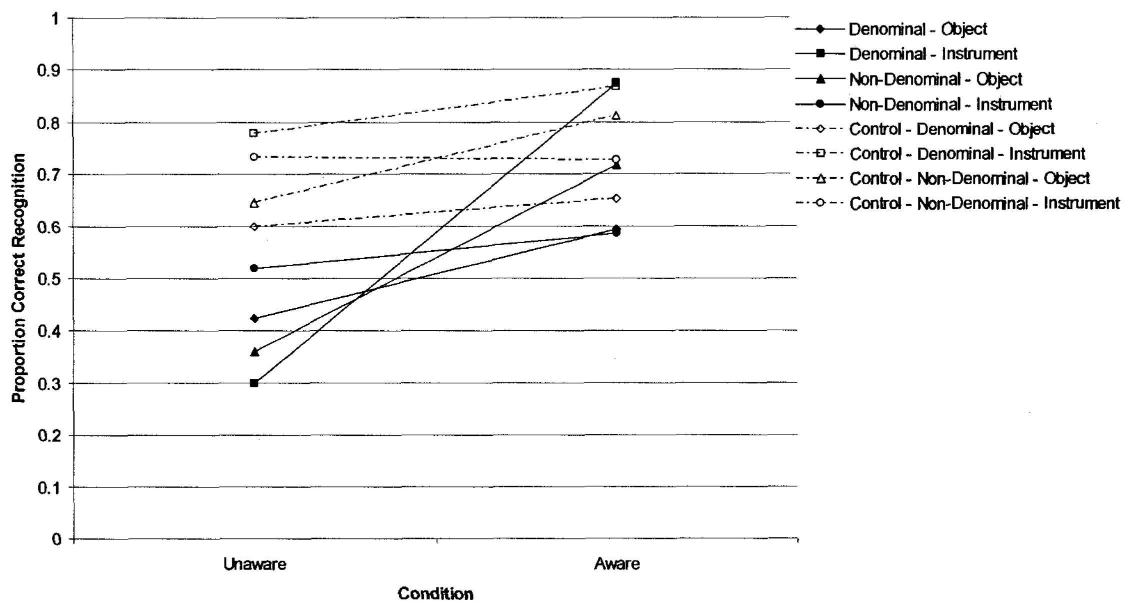
Study	Measure	Factor	Analysis	p	η^2
No	Correct	Change	F1(2,10)= 18.0	.001 *	.782
Sentence	Recognition	Detection (D)	F2(2,16)= 20.9	.001 *	.723
		Target Item (I)	F1(3,15)= 0.553	.654	.100
			F2(3,24)= 0.179	.910	.022
		D x I	F1(2,10)= 18.0	.001 *	.287
			F2(2,16)= 20.9	.001 *	.197
Sentence	Correct	Verb (V)	F1(1,18)= 1.60	.222	.082
	Recognition		F2(1,16)= 2.48	.135	.134
		Detection (D)	F1(2,36)= 18.8	.001 *	.511
			F2(2,32)= 21.5	.001 *	.574
		V x D	F1(2,36)= 0.243	.786	.013
			F2(2,32)= 0.260	.773	.016
No	Correct	Aware of	F1(1,21)= 16.1	.001 *	.435
Sentence	Recognition	Change	F2(1,17)= 30.9	.001 *	.645
		Change	F1(1,21)= 5.23	.033 *	.199
		Condition	F2(1,17)= 8.47	.010 *	.333
Sentence	Correct	Aware of	F1(1,13)= 0.858	.371	.062
	Recognition	Change	F2(1,16)= 5.14	.038 *	.243
		Change	F1(1,13)= 6.70	.022 *	.340
		Condition	F2(1,16)= 19.5	.001 *	.549

between these two factors. Bonferroni post-hoc pairwise comparisons indicated that the difference in proportion correct recognition scores lay between the undetected and detected changed items, and between the undetected and the no-change control items. More accurate recognition of the target items in the no-change control condition is not surprising considering that participants in this condition had the entire trial to attend to the target items, as compared to participants in the experimental conditions who had a limited time frame in which to encode the target items before they disappeared from the screen.

Although participants appear to have obtained higher recognition scores for the undetected and detected experimental trials of Experiment 2b, as compared to Experiment 2a, there was no associated increase in the recognition scores for the no-change control trials and change detection scores. This pattern of results suggests that either: (1) participants utilized the contents of the sentence to quickly identify and label the currently or previously attended target items, which were at times ambiguous, before they completely disappeared, and/or (2) participants responded affirmatively in the recognition task without having actually seen the target items, because they had either recalled hearing the noun referents in the accompanying sentences or having repeated the sentences in their verbal responses.

Proportion correct recognition scores in the three change conditions (i.e., no-change control, object or instrument dissolves), before and after first correct change detection were plotted in Figure 47, for Experiment 2a and 2b. For both studies because of several empty cells, proportion correct recognition scores were averaged across all four categories (i.e., denominal object and instrument; non-denominal object and instrument).

A)



B)

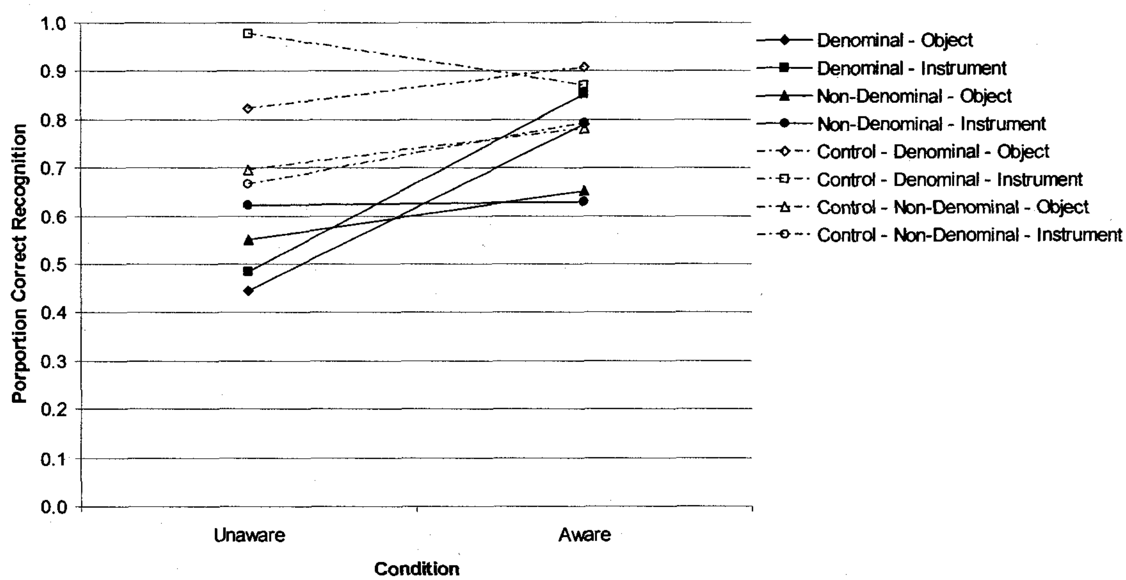


Figure 47. Proportion correct recognition of target items in the three change conditions, across verb classes, before (unaware) and after (aware) first correct change detection, for A) Experiment 2a and B) Experiment 2b.

As shown in the third panel of Table 12, for Experiment 2a the analysis revealed a statistically significant main effect of change awareness and change condition, and no significant interaction between these two factors, $\underline{F1}(1,21)= 0.721$, $\underline{p}= .405$, $\underline{\eta_p}^2 = .033$; $\underline{F2}(1,17)= 2.36$, $\underline{p}= .143$, $\underline{\eta_p}^2 = .122$. Thus, once participants were no longer naïve to the possibility of item disappearance they were later more accurate at recognizing the items in the post-test recognition task. Awareness of the possibility of scene changes appears to have increased the target items degree of central interest, in particular for the denominal target items, which were now considered worth closer attention and mention.

As shown in the bottom panel of Table 12, the results were somewhat similar for Experiment 2b, for which the analysis revealed a statistically significant main effect of change awareness (for the by item analysis only) and change condition, as well as a significant interaction between these two factors, for the by item analysis only, $\underline{F1}(1,13)= 0.032$, $\underline{p}= .860$, $\underline{\eta_p}^2 = .002$; $\underline{F2}(1,16)= 4.41$, $\underline{p}= .052$, $\underline{\eta_p}^2 = .216$. Hence, similar to Experiment 2a, participants performed better in the recognition task in the no-change control condition than in the experimental trials (i.e., target items gradually dissolved), because the former had the entire trial to attend to the target items. Furthermore, once participants became aware that scene changes could occur recognition for the target items increased, in particular for the experimental trials. One possibility is that after the first correct change detection these target items, in particular the denominal object and instrument items, became more central interest areas, which then resulted in earlier saccades to the items before they disappeared.

2.4.2.6 Links between each Overt Measure

In summary, to examine the relationships between all of these overt measures Phi correlations were calculated between: (1) correct change detection for each target item and their corresponding mention, identification, and recognition scores; (2) mention of the target items in the verbal responses and their corresponding identification and recognition scores; and (3) correct identification of the target items and their corresponding recognition scores. As shown in Tables 13 to 16, across all four categories of target items there were statistically significant positive associations between change detection rates and mention of the target items in participants' verbal responses. This is not surprising considering that correct change detection, unbeknownst to participants, was determined by their verbal descriptions of the scene. The correlations, however, were not perfect because in a few trials, participants had mentioned the target items, yet did not notice their disappearance.

On the other hand, the degree of accurate identification of the target items, for the most part, was not found to be significantly associated with change detection performance. The two experiments differed with regards to the degree of association between correct recognition and identification of the target items. In experiment 2a there was no significant relationship between these two measures, whereas a positive relationship was indicated in Experiment 2b. This appears to lend further support to the hypothesis that recall of the uttered sentence in Experiment 2b, later resulted in more accurate performance on the post-test recognition and identification tasks. The performance rates obtained in these tasks may have, therefore, overestimated participants' actual visual memory of the target items, as well as their proper identification during the experiment. Finally, across all four categories there were also statistically significant positive associations between subsequent recognition of the target

Table 13. Phi coefficients for degree of associations between proportion correct change detection, proportion mentioned in verbal response, proportion correct identification, and proportion correct recognition, in the denominal object change conditions, for A) Experiment 2a and B) Experiment 2b.

A)

Denominal Verb / Object Disappears	*Prop. Correct Change Detection*	*Proportion Mentioned*	*Prop. Correct Identification*	*Prop. Correct Recognition*
Prop. Correct Change Detection	---	0.949 * $p < .001$	- 0.053 $p = 0.631$	0.421 * $p < .001$
Proportion Mentioned		---	- 0.053 $p = 0.631$	0.653 * $p < .001$
Prop. Correct Identification			---	0.00 $p = 1.00$
Prop. Correct Recognition				---

B)

Denominal Verb / Object Disappears	*Prop. Correct Change Detection*	*Proportion Mentioned*	*Prop. Correct Identification*	*Prop. Correct Recognition*
Prop. Correct Change Detection	---	0.566 * $p < .001$	0.145 $p = 0.190$	0.446 * $p < .001$
Proportion Mentioned		---	0.133 $p = 0.230$	0.423 * $p < .001$
Prop. Correct Identification			---	0.318 * $p = .004$
Prop. Correct Recognition				---

Table 14. Phi coefficients for degree of associations between proportion correct change detection, proportion mentioned in verbal response, proportion correct identification, and proportion correct recognition, in the denominal instrument change conditions, for A) Experiment 2a and B) Experiment 2b.

A)

Denominal Verb / Instrument Disappears	*Prop. Correct Change Detection.*	*Proportion Mentioned*	*Prop. Correct Identification*	*Prop. Correct Recognition*
Prop. Correct Change Detection	---	0.894 * p < .001	0.037 p = 0.739	0.678 * p < .001
Proportion Mentioned		---	0.146 p = 0.189	0.722 * p < .001
Prop. Correct Identification			---	- 0.007 p = 0.949
Prop. Correct Recognition				---

B)

Denominal Verb / Instrument Disappears	*Prop. Correct Change Detection.*	*Proportion Mentioned*	*Prop. Correct Identification*	*Prop. Correct Recognition*
Prop. Correct Change Detection	---	0.566 * p < .001	0.253 * p = 0.023	0.471 * p < .001
Proportion Mentioned		---	0.210 p = 0.059	0.317 * p = .004
Prop. Correct Identification			---	0.406 * p < .001
Prop. Correct Recognition				---

Table 15. Phi coefficients for degree of associations between proportion correct change detection, proportion mentioned in verbal response, proportion correct identification, and proportion correct recognition, in the non-denominal object change conditions, for A) Experiment 2a and B) Experiment 2b.

A)

Non-Denominal Verb / Object Disappears	*Prop. Correct Change Detection*	*Proportion Mentioned*	*Prop. Correct Identification*	*Prop. Correct Recognition*
Prop. Correct Change Detection	---	0.901 * p < .001	0.079 p = 0.477	0.447 * p < .001
Proportion Mentioned		---	0.097 p = 0.381	0.539 * P < .001
Prop. Correct Identification			---	0.239 * p = 0.032
Prop. Correct Recognition				---

B)

Non-Denominal Verb / Object Disappears	*Prop. Correct Change Detection*	*Proportion Mentioned*	*Prop. Correct Identification*	*Prop. Correct Recognition*
Prop. Correct Change Detection	---	0.588 * p < .001	0.106 p = 0.342	0.323 * p = .004
Proportion Mentioned		---	0.117 p = 0.292	0.433 * P < .001
Prop. Correct Identification			---	0.341 * p = 0.002
Prop. Correct Recognition				---

Table 16. Phi coefficients for degree of associations between proportion correct change detection, proportion mentioned in verbal response, proportion correct identification, and proportion correct recognition, in the non-denominal instrument change conditions, for A) Experiment 2a and B) Experiment 2b.

A)

Non-Denominal Verb / Instrument Disappears	*Prop. Correct Change Detection*	*Proportion Mentioned*	*Prop. Correct Identification*	*Prop. Correct Recognition*
Prop. Correct Change Detection	---	0.840 * p < .001	- 0.138 p = 0.216	0.252 * p = 0.023
Proportion Mentioned		---	- 0.121 p = 0.275	0.320 * p = .004
Prop. Correct Identification			---	- 0.031 p = 0.781
Prop. Correct Recognition				---

B)

Non-Denominal Verb / Instrument Disappears	*Prop. Correct Change Detection*	*Proportion Mentioned*	*Prop. Correct Identification*	*Prop. Correct Recognition*
Prop. Correct Change Detection	---	0.590 * p < .001	0.088 p = 0.427	0.250 * p = 0.047
Proportion Mentioned		---	0.239 * p = 0.031	0.311 * p = .005
Prop. Correct Identification			---	0.189 p = 0.088
Prop. Correct Recognition				---

items and both correct change detection and mention of the target items. These results suggest that the more likely participants mentioned the target items in their verbal responses, the more likely it was because they had correctly detected the scene changes, and subsequently the more likely they were of accurately recognizing the target items in the recognition task.

Although, these overt measures provided us with valuable information on a number of the factors which resulted in successful change detection, an analysis of the eye movement data will hopefully provide additional information regarding the change blindness phenomenon and the nature of our visual representations. An attempt will be made at answering such questions as: (1) Did correct change detection require prior fixation of the target items, and did prior fixation guarantee correct change detection? (2) Did correct change detection result from fixating the target items while they were gradually changing, or from re-fixating the target items after they had disappeared, and did fixations during or after the change guarantee correct change detection? (3) Are the higher recognition rates for participants who detected the scene changes, the consequence or cause of successful change detection? (4) Once participants were no longer naïve about the possibility of scene changes, did this knowledge affect viewing behavior, and if so, how? Last but not least, with respect to the central issue addressed in this paper (i.e., the nature of the visual-linguistic interaction), the eye movement data will be analyzed to examine the effects of the semantic-selection properties of the different classes of verbs in guiding visual attention toward the referents of their grammatical objects and instruments during sentence comprehension.

The EyeLink Data Viewer software was used to produce a scan path of the saccades and fixations across the scene, as illustrated in Figure 48. Only scan paths from 21 of the original 27 participants could be used for Experiment 2a and 2b, because of some severely degraded eye movement data. This degradation tended to occur gradually over the course of the experiment, suggesting that it was a consequence of eye tracker slippage. This tends to occur with greater frequency when the head is not kept still throughout the experiment, as was the case with these studies in which participants were required to talk for 15 seconds after each trial. Several sets of analyses were conducted on the eye movement data, as described below.

Percent Fixation Durations of the Target Items. The fixation durations for each target item were added and averaged across the three change conditions (i.e., no-change, object or instrument items disappear) and the two verb classes (i.e., denominal and non-denominal verbs), then transformed into a percentage of total trial duration. This measure was examined to determine whether there were any significant differences in the length of time participants fixated the target items in the visual scenes. In addition, percent fixation durations were examined relative to onset of disappearance (i.e., before, during and after the scene change) and compared between participants who successfully detected and those who were blind to the scene changes; as well as between incidental and intentional change detection trials. The hypothesis was that participants who successfully detect the scene changes will spend more time fixating the changing items, in particular while they are gradually changing, than those who will not detect the scene changes. Moreover, the prediction was that if participants have a stable and accurate visual representation of the scene and eye movements are closely time-locked to the utterances in the sentences, then participants will fixate the target items after their disappearance

Figure 48. Sample scene and scan pattern of a participant in the denominal verb (i.e. beat) – object (i.e. eggs) change condition. The red lines and circles represent saccades and fixations made before the change, the green lines and circles during the gradual change, and the blue lines and circles represent saccades and fixations made after the change. The diameters of the circles, which correspond to fixations, are proportional to their duration, indicated by the numbers located at their periphery. In this case, the participant was no longer naïve to the purposes of the study, and had correctly detected the gradual dissolving of the target object. The participant had not fixated the target object prior to the change, but he/she did fixate the target object several times until it completely dissolved.

(when the critical linguistic information is encountered). Finally, it was anticipated that once participants are no longer naïve to the possibility of scene changes, they will spend more time fixating the target items.

Number of Saccades to the Target Items. The number of saccades made to each target item were summed and then averaged across the three change conditions and the two verb classes. This measure was studied to determine whether there were any significant differences in the number of saccades participants made to the target items in the visual scenes. Furthermore, the mean number of saccades were examined relative to the onset of the disappearance and compared between participants who successfully detected the scene changes and those who were blind to the scene changes; as well as between incidental and intentional change detection trials. The prediction was that participants that successfully detect the scene changes will make more saccades to the changing items, in particular while they are gradually changing, than those who will not detect the scene changes. Moreover, the expectation was that if participants have a stable and accurate visual representation of the scene and eye movements are closely time-locked to the utterances in the sentences, then participants will initiate saccades towards the target items after their disappearance. Lastly, it was anticipated that once participants are no longer naïve to the possibility of scene changes, they will initiate more saccades to the target items.

Sequence of Eye Movements to the Target Items: The percentage of trials in which participants executed the eight possible eye movement sequences (i.e., only fixated target item before, during or after the change; before and during the change; before and after the change; during and after the change; before, during and after the change; and never fixated the target item) were examined relative to accurate change detection. This measure was examined to determine whether timing of the eye movements towards the target items, relative to disappearance onset and offset, was of significant importance in accurate change detection. The hypothesis was that correct change detection will increase as participants spend more time fixating the target items before they completely disappear, and in particular while they are gradually changing.

Cumulative Saccades to the Target Items. The cumulative number of saccades that were initiated towards the target items during each 50 ms interval following the onset of the movie was calculated. For each critical point in the sentence (i.e., verb, object and instrument onset), means for each condition (i.e., denominal object and instrument, non-denominal object and instrument) were obtained by dividing the cumulative number of saccades by the total number of trials. This measure was examined to determine whether verb-guided eye movements would be closely time-locked to the utterance of the verb, in particular before the onset of the object and instrument nouns in the sentences. The prediction was that the mean number of cumulative saccades to the target items will increase as the sentences unfold, because processing of the verbs and their grammatical objects and instruments occurs incrementally, thus, providing participants with more restrictive information as time progresses.

First Post-Verb Saccades to the Target Items. The delays in initiating the first saccades towards the referents of the object and instrument nouns, after hearing the verb, were obtained and then compared against two other time points in the sentence: the object and instrument noun onsets. This measure was studied in order to determine whether first saccades were affected by verb type and if they were initiated prior to the utterance of the object and instrument nouns in the sentences. Moreover, first SOTs were measured to examine whether the eye movements would be closely time-locked to the utterance of the verb, object and instrument nouns. The expectation was that if verb-specific information can constrain visual attention at the very early stages of processing, then participants will likely initiate their eye movements to these target items prior to their onset in the accompanying sentences (i.e., anticipatory eye movements).

2.4.2.7 Overall Percent Fixation Durations

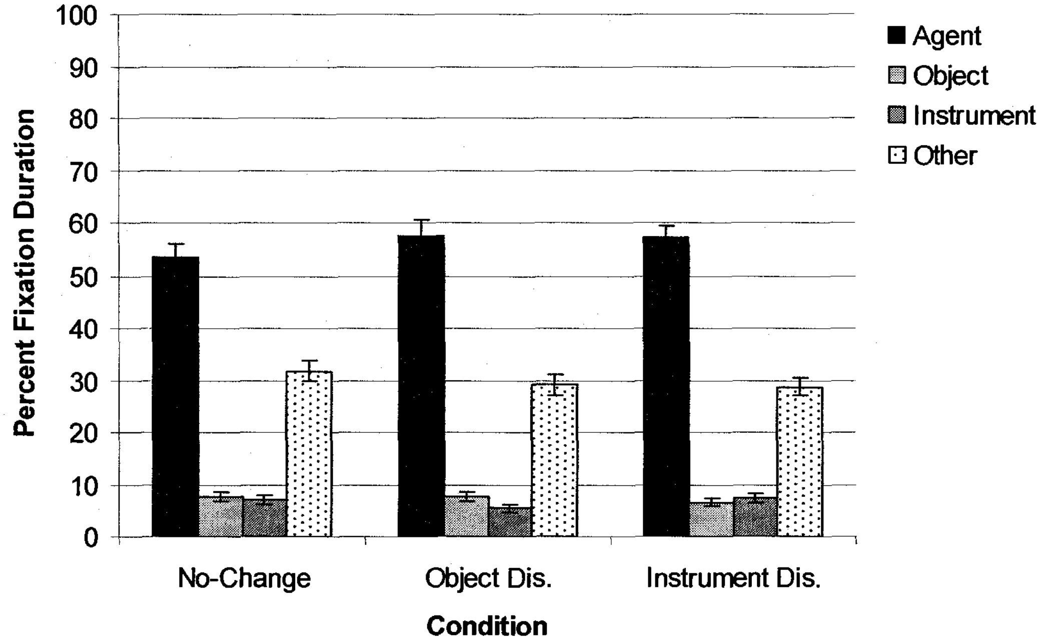
Overall percent fixation duration was defined as the total amount of time fixating each item over the total trial duration (i.e., 6 seconds). Table 17 summarizes the results of the analyses conducted on the percent fixation durations across the two experiments and Figure 49 illustrates these durations to the different scene items (i.e., agent, target object and instrument, and other) across change conditions (i.e., no-change control, object disappears, and instrument disappears). As shown in the top panel of Table 17, for Experiment 2a the two repeated-measures ANOVAs revealed a statistically significant main effect of scene item, no significant main effect of change condition or interaction between these two factors. Bonferroni post-hoc pairwise comparisons indicated that the difference lay between the agent and target object, instrument and other scene items, as well as between the other scene items and the object and instrument items. Thus, poor change detection appears to be, in large part, the result of participants spending the majority of the trial fixating the moving agent (56.2%), rather than the target object (7.3%) and instrument (6.6%) items which were sometimes subject to change.

As indicated in the second panel of Table 17, analysis for Experiment 2b revealed similar findings, with a statistically significant main effect of scene items, no significant

Table 17. Summary of analyses conducted on percent fixation durations, across the two experiments (without and with accompanying sentence).

Study	Measure	Factor	Analysis	p	η^2
No Sentence	Percent	Change	$F1(2,40)= 1.37$	.265	.064
	Fixation	Condition (C)	$F2(2,34)= 0.883$	.423	.049
	Duration	Scene Item (I)	$F1(3,60)= 219.2$	.001 *	.916
			$F2(3,51)= 171.6$	.001 *	.910
		C x I	$F1(6,120)= 1.50$	.227	.070
			$F2(6,102)= 0.941$	.428	.052
Sentence	Percent	Change	$F1(2,40)= 0.06$	.942	.003
	Fixation	Condition (C)	$F2(2,34)= 1.92$	.163	.101
	Duration	Scene Item (I)	$F1(3,60)= 171.9$	.001 *	.896
			$F2(3,51)= 276.5$	.001 *	.942
		C x I	$F1(6,120)= 1.16$	.334	.055
			$F2(6,102)= 1.42$	.214	.077
No Sentence	Percent	Verbal Cue	$F1(1,40)= 0.380$	.541	.009
vs. Sentence	Fixation		$F2(1,34)= 0.035$	.085	.001
	Duration				

A)



B)

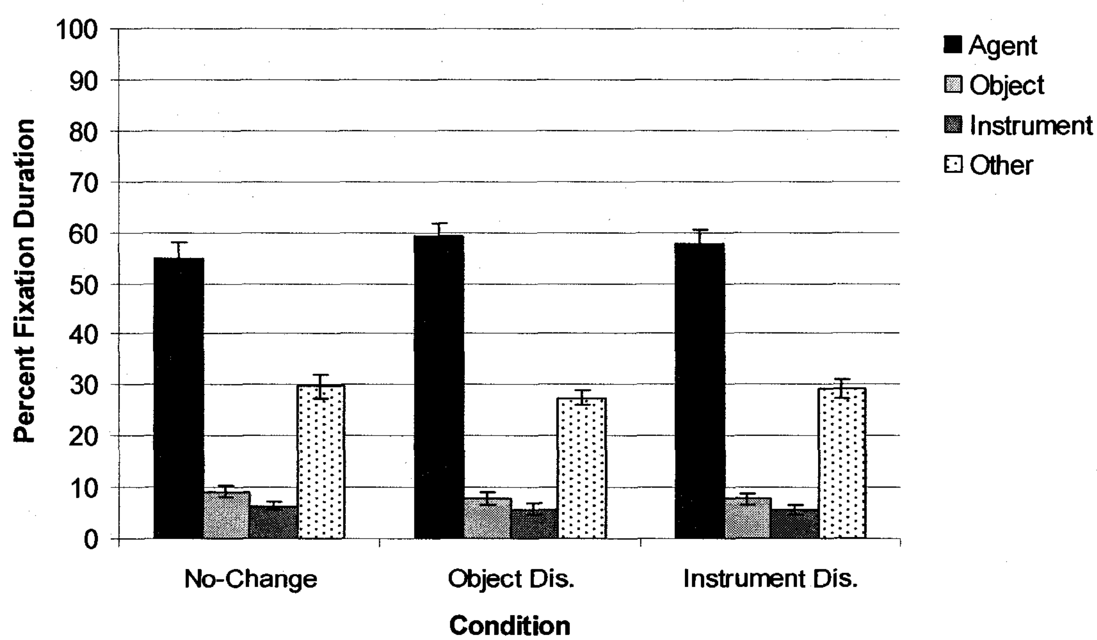


Figure 49. Percent fixation durations for each scene item, across change conditions, for A) Experiment 2a and B) Experiment 2b.

main effect of change condition or significant interaction between these two factors. Bonferroni post-hoc pairwise comparisons indicate that the difference lay between the agent and object, instrument and other scene items, between the other scene items and the object and instrument items, as well as between the object and instrument items for the by subjects analysis. Thus, similar to Experiment 2a, participants spent most of the trial fixating the moving agent (54.74%), rather than the target object (9.22%) and instrument (6.32%) items which were sometimes subject to change.

The percent fixation durations were then compared across the two studies and, as shown in the bottom panel of Table 17, the analysis revealed no statistically significant main effect of sentential context, as well as no significant interaction between sentence and scene items $\underline{F1}(3,120)= 0.205$, $\underline{p}= .892$, $\underline{\eta_p}^2 = .005$; $\underline{F2}(3,102)= 0.241$, $\underline{p}= .868$, $\underline{\eta_p}^2 = .007$, or between sentence and change conditions $\underline{F1}(2,80)= 0.769$, $\underline{p}= .467$, $\underline{\eta_p}^2 = .008$; $\underline{F2}(2,68)= 0.554$, $\underline{p}= .577$, $\underline{\eta_p}^2 = .016$. Overall, contrary to expectations, participants in Experiment 2b did not spend more time fixating the target objects and items which were directly referred to in the accompanying sentences. This pattern of results may explain why change detection performance did not increase in Experiment 2b. If participants had made eye movements to the referred items as they were being uttered in the sentence, then a return to the location of the previously present target item would have resulted in greater change detection.

2.4.2.8 Percent Fixation Durations Relative to Disappearance Onset

What is crucial to successful change detection, however, may not be the total duration of time spent fixating the target items, but rather *when* during the change process time was spent attending the target items. Table 18 summarizes the results of the analyses conducted on the percent fixation durations relative to disappearance onset (i.e., before, during and after item disappearance) across the two experiments and Figure 50 illustrates these durations across verb types and target items, for participants who successfully and unsuccessfully detected the scene changes. For Experiment 2a, because of multiple empty cells, percent fixation durations were averaged across all four change conditions.

As shown in the top panel of Table 18, the analysis revealed a statistically significant main effect of fixation timing, change detection, and a significant interaction between these two factors. Bonferroni post-hoc pairwise comparisons indicated that the percent fixation duration difference lay between the before and during, and before and after time sections. Hence, participants who successfully noticed the gradual dissolving target items had spent more time fixating the target items throughout the trial, as compared to participants who did not notice the scene change. Although change detectors spent slightly more time fixating the target items prior to the change, it appears that fixation of the target item while it changed, as well as after it had completely disappeared from the screen, was what resulted in successful change detection.

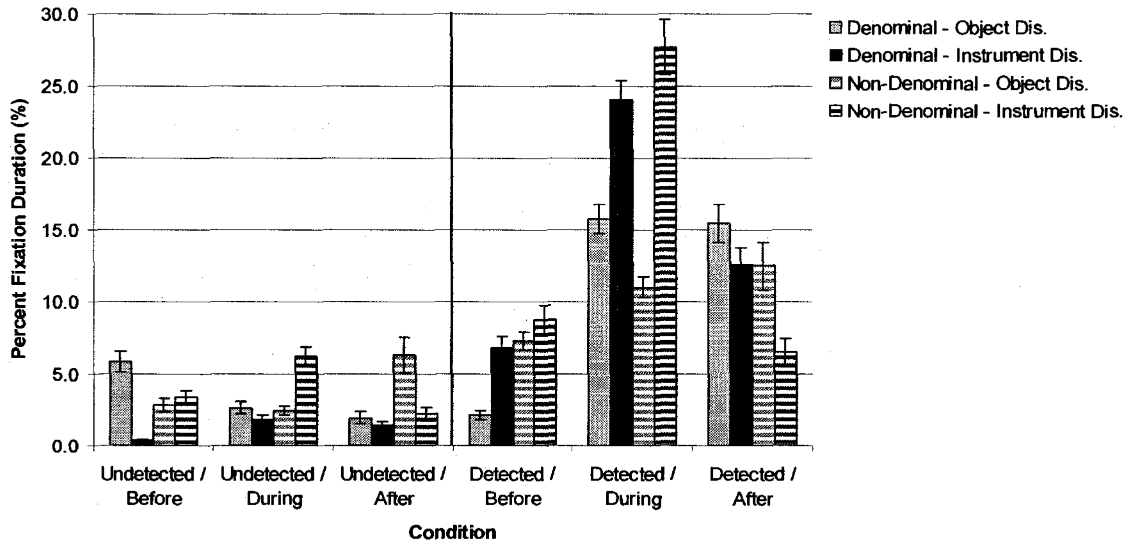
As shown in the second panel of Table 18, similar results were obtained in Experiment 2b, with no statistically significant main effect of verb type, but a statistically significant main effect of fixation timing, change detection, and a significant interaction between these two factors. The proportion of fixation durations relative to disappearance onset were then compared across the two studies and, as shown in the third panel of Table 18, the analysis revealed no significant main effect of sentential context, as well as no significant interaction between sentence and fixation timing, $F1(2,54) = 0.269$, $p = .765$,

Table 18. Summary of analyses conducted on percent fixation duration, relative to disappearance onset, across the two experiments (without and with accompanying sentence).

Study	Measure	Factor	Analysis	p	η^2
No Sentence	Percent Fixation	Detection (D)	F1(1,20)= 60.29	.001 *	.766
	Duration		F2(1,17)= 93.24	.001 *	.846
		Timing (T)	F1(2,40)= 11.78	.001 *	.371
			F2(2,34)= 10.17	.001 *	.374
		D x T	F1(2,40)= 9.87	.001 *	.330
			F2(2,34)= 10.12	.001 *	.373
Sentence	Percent Fixation	Verb (V)	F1(1,13)= 1.15	.304	.081
	Duration		F2(1,16)= 0.000	.994	.000
		Detection (D)	F1(1,13)= 13.66	.003 *	.512
			F2(1,15)= 74.65	.001 *	.823
		Timing (T)	F1(2,26)= 7.34	.003 *	.361
			F2(2,30)= 6.32	.005 *	.283
		D x T	F1(2,26)= 7.09	.003 *	.353
			F2(2,30)= 5.60	.008 *	.259
No Sentence	Percent Fixation	Verbal Cue	F1(1,27)= 0.537	.470	.019
vs. Sentence	Duration		F2(1,31)= 0.438	.513	.014
No Sentence	Percent Fixation	Aware of	F1(1,13)= 14.84	.002 *	.533
	Duration	Change	F2(1,17)= 10.79	.004 *	.388
		Change	F1(1,13)= 5.05	.043 *	.280
		Condition	F2(1,17)= 9.19	.008 *	.351

Sentence	Percent Fixation	Aware of	$F1(1,10)= 11.15$	.008 *	.527
	Duration	Change	$F2(1,17)= 50.15$	.000 *	.747
		Change	$F1(1,10)= 15.09$	.003 *	.601
		Condition	$F2(1,17)= 8.59$	.009 *	.336

A)



B)

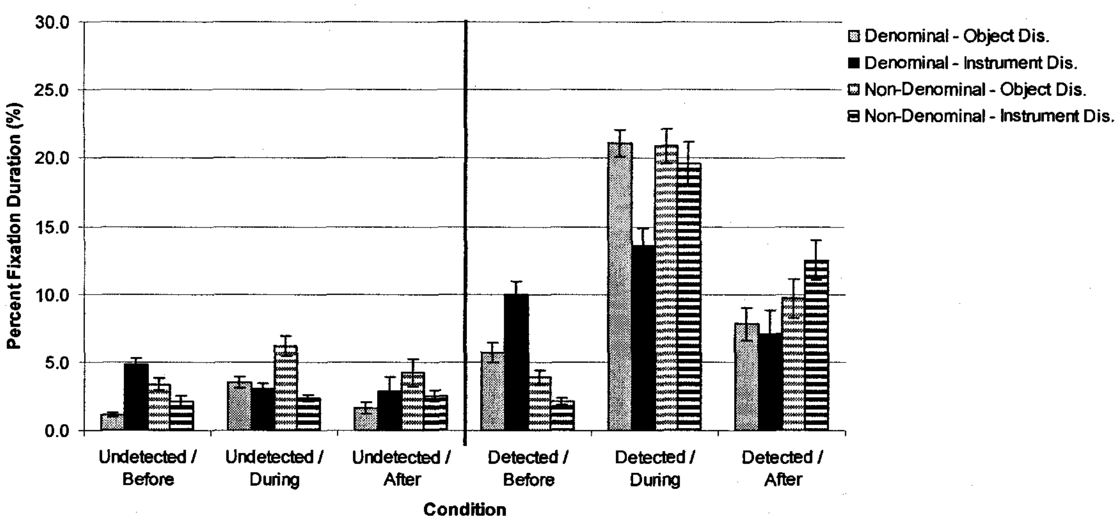


Figure 50. Percent fixation durations relative to disappearance onset (before, during and after disappearance), across change conditions, for successful (detected) and unsuccessful (undetected) detection of changed items, in A) Experiment 2a and B) Experiment 2b.

η_p^2 = .010; F2(2,64)= 0.708, p= .496, η_p^2 = .021, between sentence and change detection, except for the items analysis, F1(1,27)= 1.97, p= .172, η_p^2 = .085; F2(1,32)= 4.64, p= .039, η_p^2 = .127, and between sentence and verb type, F1(1,27)= 0.524, p= .474, η_p^2 = .019; F2(1,32)= 0.249, p= .621, η_p^2 = .008. Contrary to expectations the sentence context, in particular the degree of selectional restrictions of the two classes of verbs, did not result in different fixation durations relative to disappearance onset. Hence, similar to Experiment 2a, fixation of the target items while they gradually changed, as well as after they had completely disappeared from the screen, resulted in accurate change detection.

Although sentence context does not appear to have facilitated change detection, it did significantly direct visual attention to the target items in the scene. Figure 51 illustrates the percent fixation durations for the no-change control trials, before verb onset and after verb onset, across verb types and target items. The analysis revealed a statistically significant main effect of sentence point, F1(1,20)= 4.39, p= .049, η_p^2 = .180; F2(1,16)= 8.89, p= .009, η_p^2 = .357, and target items, F1(1,20)= 8.05, p= .010, η_p^2 = .287; F2(1,16)= 8.61, p= .010, η_p^2 = .350, yet no main effect of verb type, F1(1,20)= 0.024, p= .878, η_p^2 = .001; F2(1,16)= 0.315, p= .582, η_p^2 = .011, and no significant interaction between any of these factors. Since participants' visual attention in the no-change control trials could not be captured by dissolving target items, the increased fixation time durations obtained after verb onset appear to be the result of visual attention being driven by the contents of the accompanying sentence.

These results, along with the findings reported in previous sections, suggest that the lack of increase in change detection rates in Experiment 2b were not the result of participants having ignored the contents of the sentence. Rather, one possibility is that

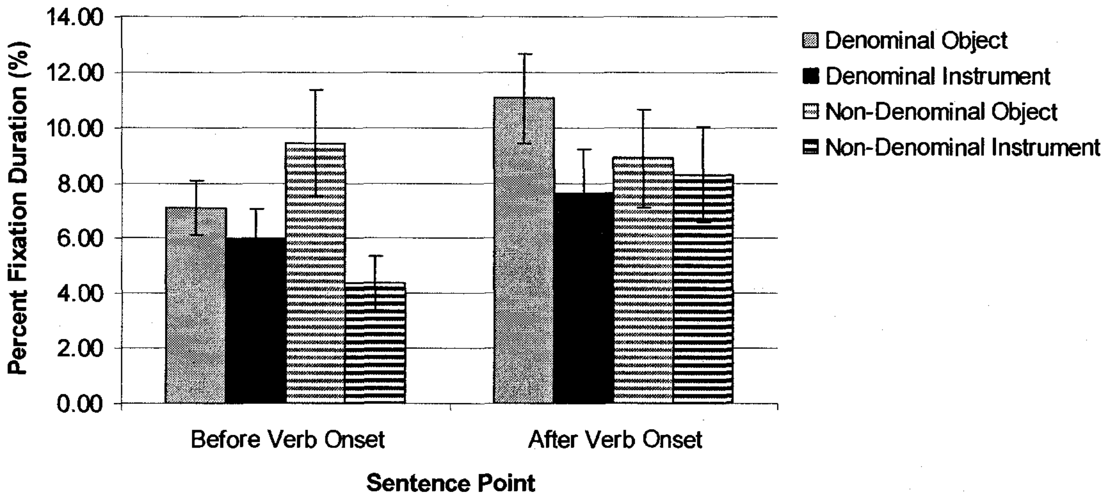
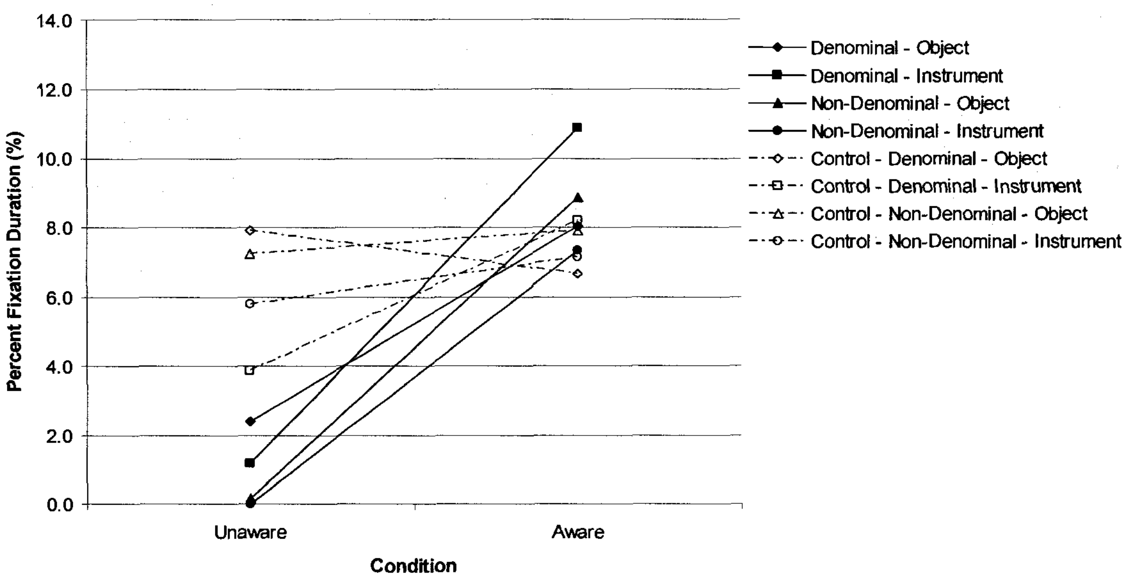


Figure 51. Percent fixation durations of target items before and after verb onset, across verb types and target items.

contrary to rapid scene changes, for which re-fixation of the target item's previously held location produces efficient change detection, gradual scene changes may instead require prolonged viewing of the target items as they are slowly changing. Quick fixations of the target items, as were produced in the undetected trials, may miss the more subtle change signals produced by gradual scene changes. The slight decrease in fixation durations during and after the scene changes in Experiment 2b, despite similar detection rates across the two studies, suggest that sentential context helped increase the efficiency of participants viewing behaviour. The contents of the sentences seem to have helped participants identify and label the items in the visual display, some of which may have been ambiguous (e.g., non-denominal instrument items). This in turn, appears to have decreased the amount of time required to fixate or re-fixate the target items, as well as increase the effectiveness of peripheral vision.

Finally, similar to the analyses reported above, it was hypothesized that once participants become aware that items in the scenes could dissolve or disappear, they would spend more time fixating the target items in the scene. Percent fixation durations in the three change conditions, before and after the first correct change detection, were plotted in Figure 52, for Experiment 2a and 2b. For both experiments, the percent fixation durations were averaged across all four categories (i.e., denominal object and instrument disappears, and non-denominal object and instrument disappears). As shown in the fourth panel of Table 18, for Experiment 2a the analysis revealed a statistically significant main effect of change awareness, change condition, and a significant interaction between these two factors, $\underline{F1}(1,13)= 15.06$, $\underline{p}= .002$, $\eta_p^2= .537$; $\underline{F2}(1,17)= 10.91$, $\underline{p}= .004$, $\eta_p^2= .391$. Thus, participants spent more time fixating the target items in

A)



B)

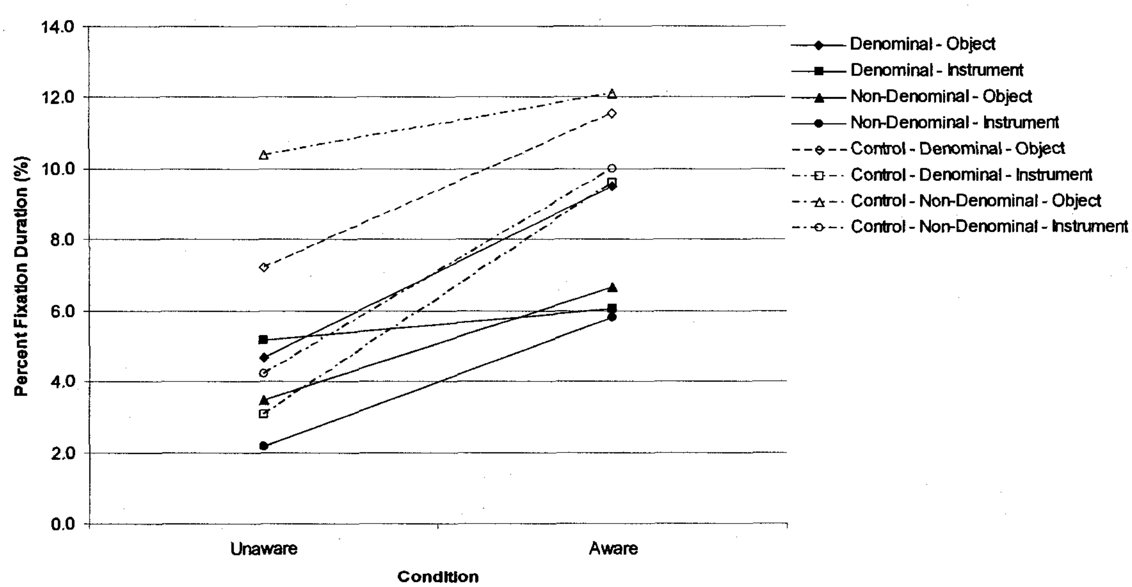


Figure 52. Percent fixation durations of target items in the three change conditions, before (unaware) and after (aware) first correct change detection, for A) Experiment 2a and B) Experiment 2b.

the no-change control condition than the change conditions (i.e., objects or instruments dissolve), but only when unaware that changes to items within the scene could occur. This is not surprising considering that in the no-change control condition the target items are displayed for the full length of the trial and, therefore, provide participants with more time to fixate these items.

As indicated in bottom panel of Table 18, analysis for Experiment 2b revealed similar findings, with a statistically significant main effect of change awareness, change condition, yet no significant interaction between these two factors, $\underline{F1}(1,10) = 0.186$, $\underline{p} = .676$, $\underline{\eta}_p^2 = .018$; $\underline{F2}(1,17) = 0.044$, $\underline{p} = .836$, $\underline{\eta}_p^2 = .003$. Combined with the results from the previous section, the findings suggest that change blindness occurred when participants either failed to fixate or did not fixate the target items long enough before the target items completely dissolved from the screen. Awareness of the detection task may have increased the saliency of the target items, as well as caused participants to become hyper vigilant to any change signal, which in turn resulted in quicker eye movements to the target items, longer fixations to the changing items and, thus, greater change detection performance.

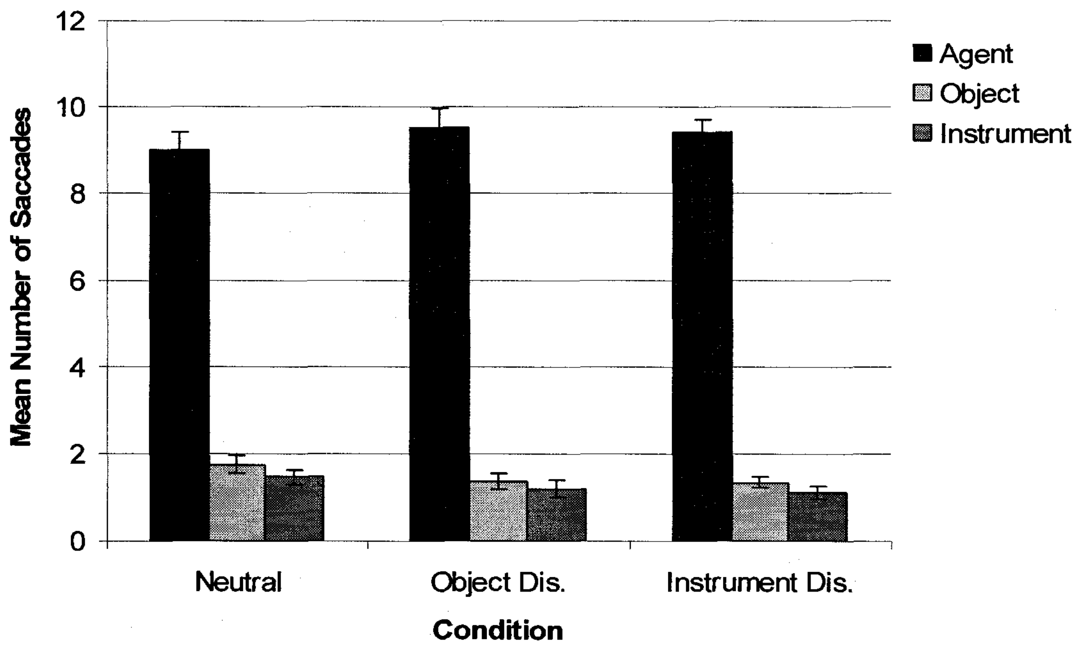
2.4.2.9 Overall Number of Saccades

A similar pattern of results was found when conducting an analysis on the overall number of saccades, that is the sum of all saccades made to each scene item over a complete trial. Table 19 summarizes the results of the analyses conducted on the mean number of saccades across the two experiments and Figure 53 illustrates these means for each scene item. As shown in the top panel of Table 19, for Experiment 2a the analysis revealed a statistically significant main effect of scene item, no significant main effect of

Table 19. Summary of analyses conducted on mean number of saccades, across the two experiments (without and with accompanying sentence).

Study	Measure	Factor	Analysis	p	η^2
No Sentence	Mean Number	Change	$F1(2,40)= 0.455$	.637	.022
	of Saccades	Condition (C)	$F2(2,34)= 0.270$	.765	.016
		Scene Item (I)	$F1(2,40)= 469.09$	.001 *	.959
			$F2(2,34)= 349.15$	.001 *	.954
		C x I	$F1(4,80)= 2.21$	.105	.099
			$F2(4,68)= 1.95$	.138	.103
Sentence	Mean Number	Change	$F1(2,40)= 0.388$	.681	.019
	of Saccades	Condition (C)	$F2(2,34)= 0.285$	.754	.014
		Scene Item (I)	$F1(2,40)= 209.70$	.001 *	.913
			$F2(2,34)= 493.71$	.001 *	.967
		C x I	$F1(4,80)= 1.70$	.175	.078
			$F2(4,68)= 2.02$	.119	.102
No Sentence	Mean Number	Verbal Cue	$F1(1,40)= 0.105$	.747	.003
vs. Sentence	of Saccades		$F2(1,34)= 0.482$	.492	.014

A)



B)

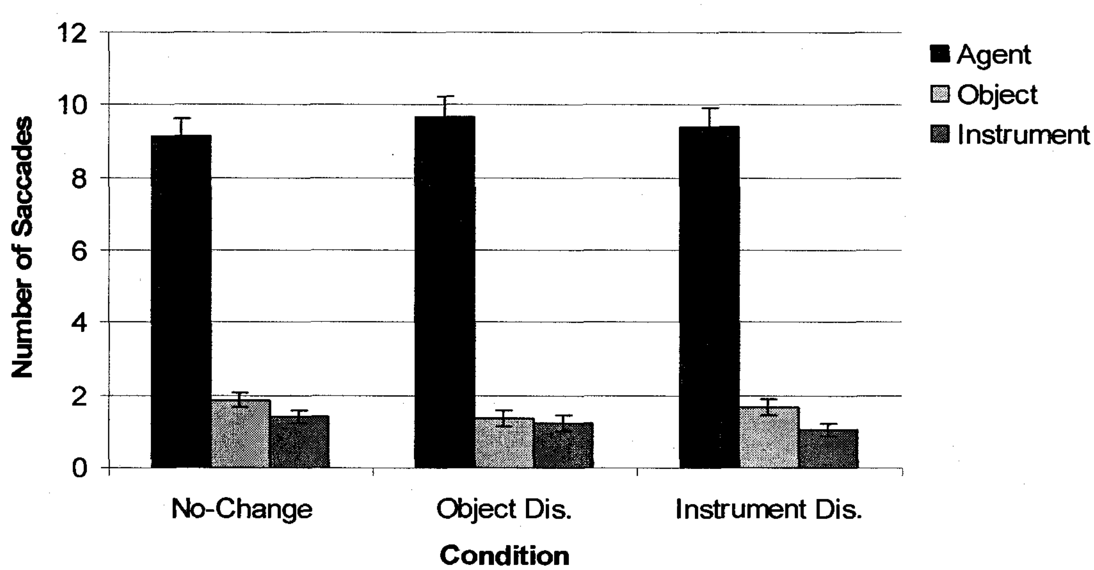


Figure 53. Average number of saccades to each scene item, across change conditions, for A) Experiment 2a and B) Experiment 2b.

change condition or interaction between these two factors. Bonferroni post-hoc pairwise comparisons indicated that the difference lay between the agent and object and instrument scene items. Combined with the percent fixation durations, these results suggest that participant's attention was for the most part directed towards the moving agent in each scene, thus diverting attention away from the changes that sometimes occurred to the object and instrument items in the scene.

As shown in the second panel of Table 19, similar results were found in Experiment 2b, for which the analysis revealed a statistically significant main effect of scene item, no statistically significant main effect of condition, or interaction between these two factors. Bonferroni post-hoc pairwise comparisons indicate that the difference lies between the agent and object and instrument scene items, and between the object and instrument items for the by subject analysis. Thus, once again the results suggest that participants made the greatest number of saccades towards the moving agent of each scene.

As shown in the bottom panel of Table 19, comparison of overall number of saccades between the two studies revealed no statistically significant effect of sentential context, and no significant interaction between sentence and scene items $F1(2,80) = 0.047$, $p = .954$, $\eta_p^2 = .001$; $F2(2,68) = 0.076$, $p = .927$, $\eta_p^2 = .002$, or between sentence and change condition $F1(2,80) = 0.009$, $p = .991$, $\eta_p^2 = .000$; $F2(2,68) = 0.046$, $p = .956$, $\eta_p^2 = .001$. Therefore, similar to the overall percent fixation duration, the accompanying sentences did not appear to have pushed participants to make more saccades to the target items, and thus become better change detectors.

2.4.2.10 Average Number of Saccades Relative to Onset of Disappearance

Table 20 summarizes the results of the analyses conducted on the mean number of saccades initiated relative to onset of disappearance (i.e., before, during and after the item disappeared) across the two experiments and Figure 54 illustrates these means across change conditions, for participants who successfully detected and those who were blind to the scene changes. For Experiment 2a, mean number of fixations were averaged across all four categories, and as shown in the top panel of Table 20, the analysis revealed a statistically significant main effect of saccade timing, change detection, and a significant interaction between these two factors. Bonferroni post-hoc pairwise comparisons indicated that the differences in mean number of saccades lay between the before and during and before and after time sections. Thus, participants who correctly detected the change initiated on average more saccades to the target items, in particular during and after the change.

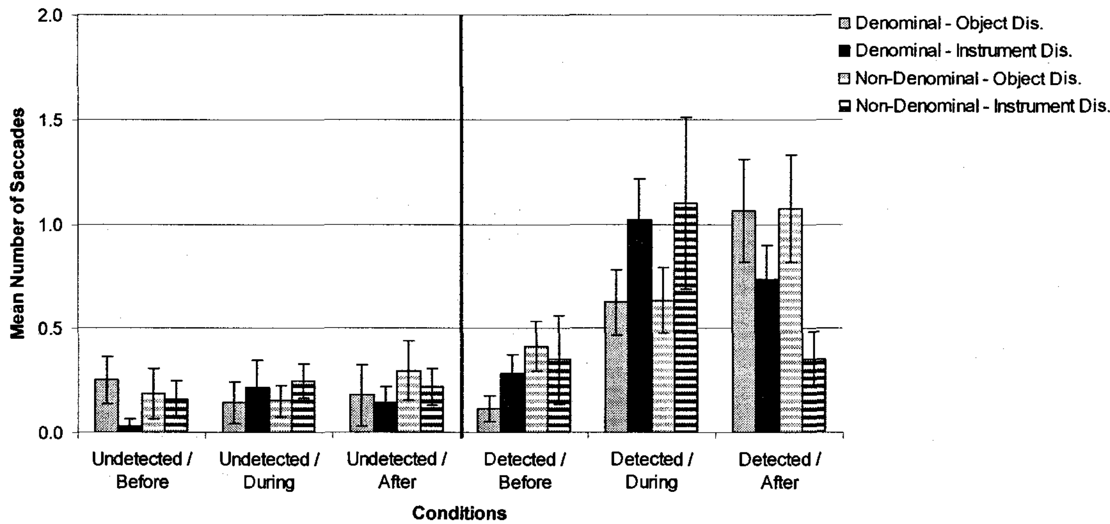
As shown in the second panel of Table 20, similar results were obtained in Experiment 2b, for which the analysis revealed no statistically significant main effect of verb, a significant main effect of saccade timing and change detection, and a significant interaction between these two factors. Bonferroni post-hoc pairwise comparisons indicated that the differences in the mean number of saccades lay between the before and during and before and after time sections. As shown in the third panel of Table 20, a comparison between the two studies for mean number of saccades indicated no statistically significant effect of sentential context, and no significant interaction between sentence and timing, $\underline{F1}(2,56)= 0.634$, $\underline{p}= .534$, $\underline{\eta_p}^2 = .022$; $\underline{F2}(2,64)= 0.733$, $\underline{p}= .485$, $\underline{\eta_p}^2 = .022$, and between sentence and change detection, $\underline{F1}(1,28)= 3.64$, $\underline{p}= .067$, $\underline{\eta_p}^2 = .115$;

Table 20. Summary of analyses conducted on mean number of saccades, relative to disappearance onset, across the two experiments (without and with accompanying sentence).

Study	Measure	Factor	Analysis	p	η^2
No Sentence	Mean Number	Detection (D)	$F1(1,20)= 45.62$	.001 *	.695
	of Saccades		$F2(1,17)= 121.57$	.001 *	.877
		Timing (T)	$F1(2,40)= 15.07$	.001 *	.430
			$F2(2,34)= 9.90$	.001 *	.368
		D x T	$F1(2,40)= 8.94$	.001 *	.309
			$F2(2,34)= 9.50$	.001 *	.359
Sentence	Mean Number	Verb (V)	$F1(1,13)= 1.33$	.269	.093
	of Saccades		$F2(1,16)= 0.092$	.766	.006
		Detection (D)	$F1(1,13)= 13.26$	.003 *	.505
			$F2(1,15)= 6.34$	.024 *	.297
		Timing (T)	$F1(2,26)= 0.243$	.015 *	.277
			$F2(2,30)= 0.260$	.001 *	.383
		D x T	$F1(2,26)= 4.64$	.019 *	.263
			$F2(2,30)= 22.67$	.000 *	.602
No Sentence	Mean Number	Verbal Cue	$F1(1,27)= 0.328$	.571	.012
vs. Sentence	of Saccades		$F2(1,31)= 0.103$	.750	.003
No Sentence	Mean Number	Aware of	$F1(1,13)= 5.58$	.034 *	.300
	of Saccades	Change	$F2(1,17)= 5.14$	.004 *	.388
		Change	$F1(1,13)= 13.1$	.003 *	.502
		Condition	$F2(1,17)= 19.5$	.008 *	.351

Sentence	Mean Number	Aware of	$F1(1,10)= 12.0$	.006 *	.545
	of Saccades	Change	$F2(1,17)= 46.24$	.000 *	.731
		Change	$F1(1,10)= 12.9$	.005 *	.563
		Condition	$F2(1,17)= 11.0$	.004 *	.392

A)



B)

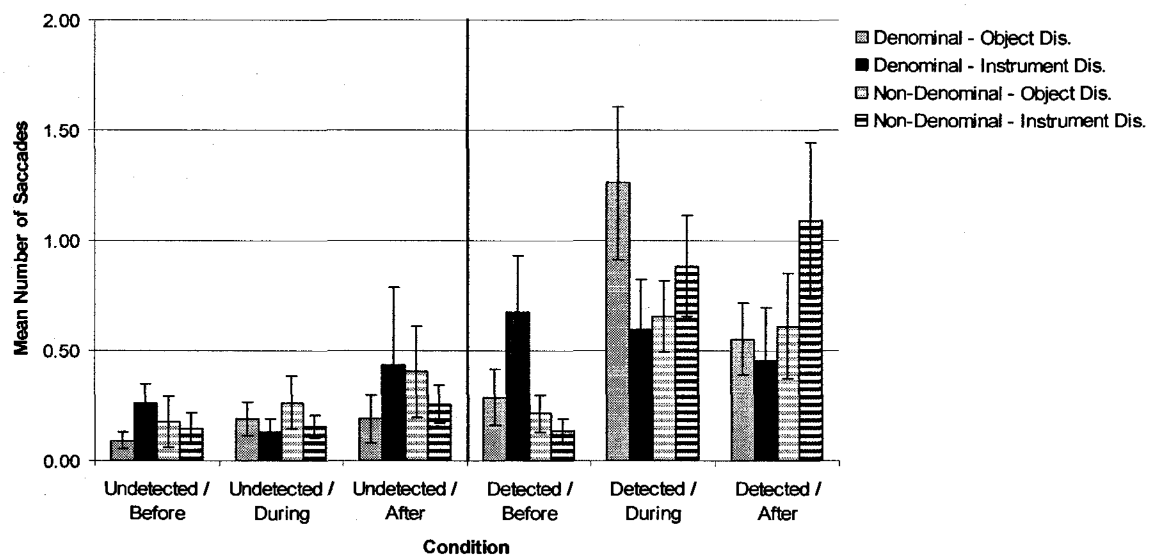


Figure 54. Mean number of saccades relative to disappearance onset (before, during and after disappearance), across change conditions, for successful (detected) and unsuccessful (undetected) detection of changed items, in A) Experiment 2a and B) Experiment 2b.

F2(1,32)= 6.81, p= .014, η_p^2 = .175. Contrary to expectations the sentence context, in particular the degree of selectional restrictions of the two classes of verbs, did not result in different mean numbers of saccades relative to disappearance onset. Hence, similar to Experiment 2a, saccades initiated to the target items while they gradually changed, as well as after they had completely disappeared from the screen, was what appears to have resulted in accurate change detection.

Yet as explained in the previous section, although sentence context does not appear to have facilitated change detection, it did significantly direct visual attention to the target items in the scene. Figure 55 illustrates the number of saccades for the no-change control trials, before verb onset and after verb onset, across verb types and target items. The analysis revealed a statistically significant main effect of sentence point, for the by subject analysis only, F1(1,20)= 25.57, p< .001, η_p^2 = .561; F2(1,16)= 2.06, p= .171, η_p^2 = .114 and target items, for the by item analysis only, F1(1,20)= 2.46, p= .132, η_p^2 = .110; F2(1,16)= 6.01, p= .026, η_p^2 = .273, yet no main effect of verb type, F1(1,20)= 0.351, p= .560, η_p^2 = .017; F2(1,16)= 0.090, p= .768, η_p^2 = .006, and no significant interaction between any of these factors. Since participants' visual attention in the no-change control trials could not be captured by dissolving target items, the increased number of saccades initiated after verb onset appear to be the result of visual attention being driven by the contents of the accompanying sentence – this is despite the fact that participants had more time to scan the visual scene prior to verb onset (i.e., 3333 ms), as compared to after verb onset (i.e. 2667 ms).

Finally, the mean number of saccades in the three change conditions was examined before and after the first correct change detection and these means were plotted

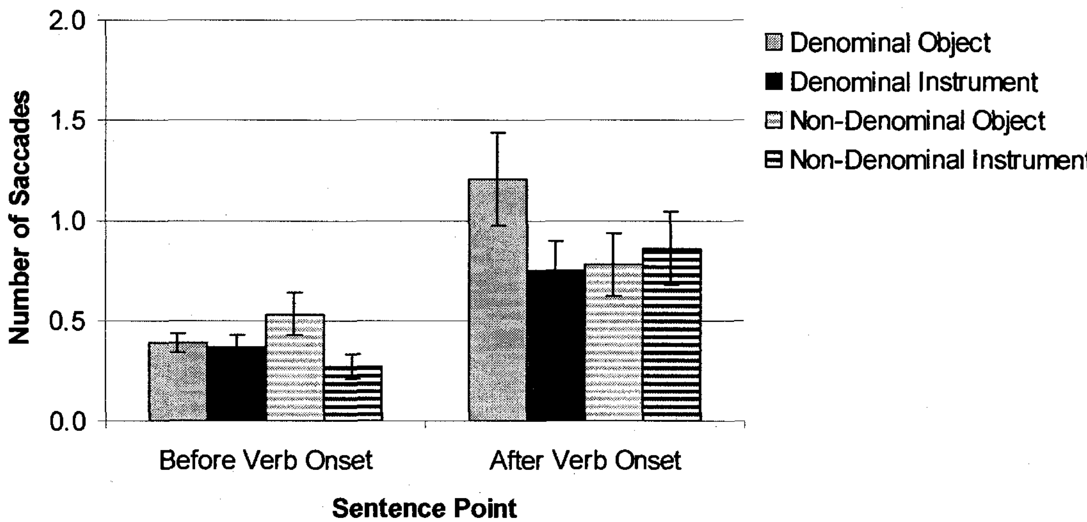


Figure 55. Number of saccades to target items before and after verb onset, across verb types and target items.

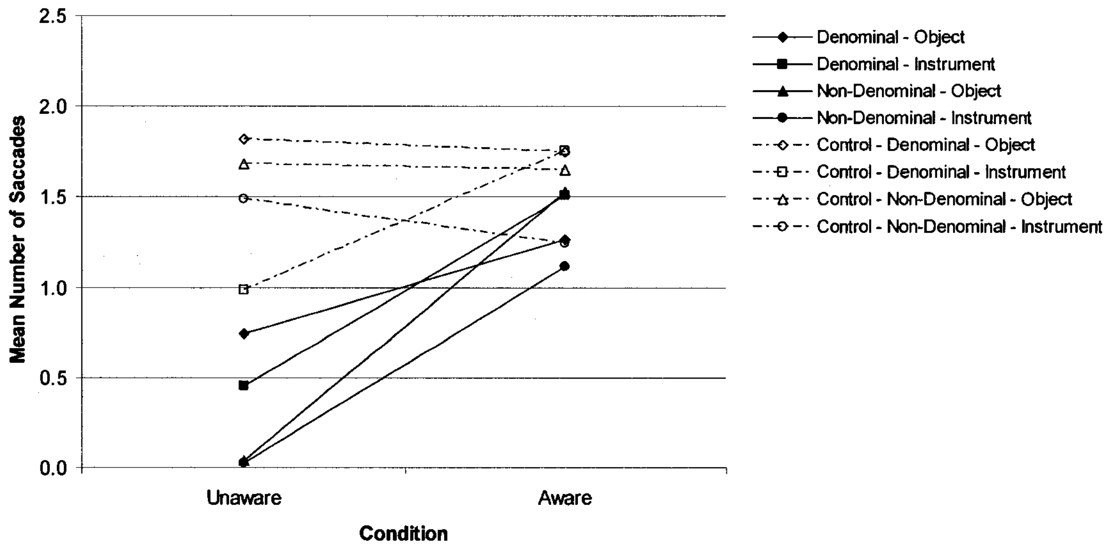
in Figure 56. For Experiment 2a, as shown in the fourth panel of Table 20, the analysis revealed a statistically significant main effect of change awareness, change condition and a significant interaction between these two factors, $F1(1,13)= 11.75$, $p= .004$, $\eta_p^2 = .475$; $F2(1,17)= 8.59$, $p= .009$, $\eta_p^2 = .336$. Thus, participants initiated on average more saccades to the target items in the no-change control condition than the change conditions (i.e., objects or instruments dissolve), but only when unaware that changes to items could occur.

As shown in the bottom panel of Table 20, similar results were found in Experiment 2b, for which the analysis revealed a statistically significant main effect of change awareness, change condition, yet there was no significant interaction between these two factors, $F1(1,10)= 1.06$, $p= .328$, $\eta_p^2 = .096$; $F2(1,17)= 0.352$, $p= .561$, $\eta_p^2 = .020$. As mentioned above, awareness of the detection task appears to have increased the saliency of the target items, as well as led participants to become hyper vigilant to any change signal, which in turn resulted in quicker eye movements to the target items, and then greater fixations and re-fixations to the changing items.

2.4.2.11 Sequence of Eye Movements

Fixations of the target items in the experimental trial, however, do not appear in and of itself sufficient to produce correct change detection. In 8.4% (n= 21) and 10.3% (n=26) of trials in Experiment 2a and 2b respectively, participants fixated the target items during or before and after the gradual change yet did not notice the scene change. Instead, the timing of the saccade towards the target item, in relation to the disappearance onset, seems to be of greater importance. For instance, if participants only fixated the target items before they disappeared, and not during or after the change, they would not

A)



B)

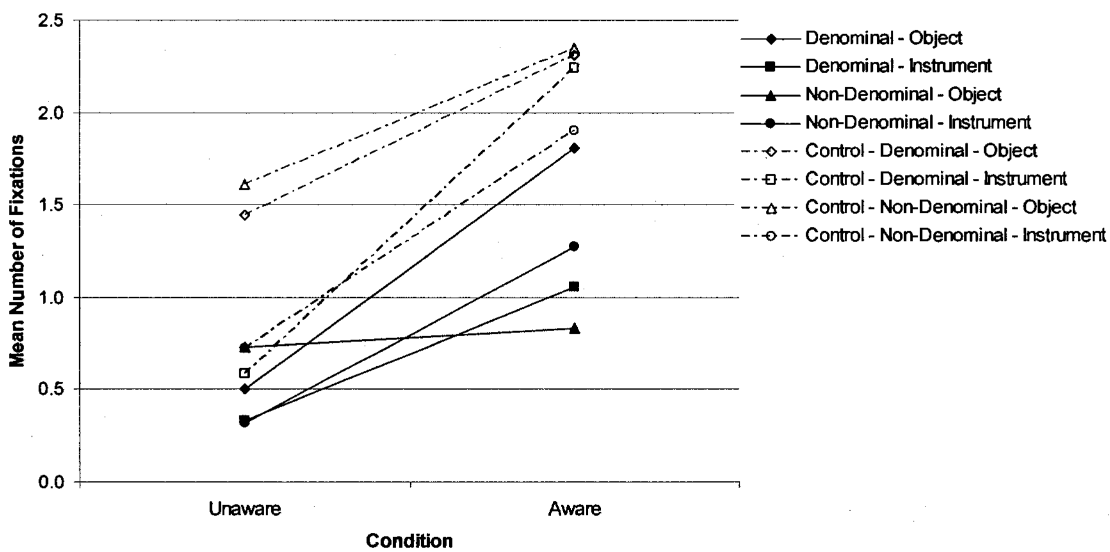


Figure 56. Mean number of fixations to the target items in the three change conditions, before (unaware) and after (aware) first correct change detection, for A) Experiment 2a and B) Experiment 2b.

be cognizant that it was no longer present. A breakdown of the percentage of trials in which participants executed the 8 possible eye movement sequences (i.e., only fixated target item before, during or after the change; before and during the change; before and after the change; during and after the change; before, during and after the change; and never fixated the target item) and the relationship with correct change detection is provided in Table 21.

Phi correlation coefficients indicated that there was a positive association between correct change detection and when participants either: 1) fixated the target item during and after the change; 2) before, during and after the change; 3) shortly after the change to the target items had occurred (generally within 200 ms of disappearance offset), for Experiment 2a only; and 4) before and during the change, for Experiment 2b only. Thus, this analysis provides further support for the proposed hypothesis that gradual change detection appears to necessitate continuous viewing of the target items until their complete disappearance. Quick fixations of the target items may miss the more subtle change signals produced by gradual scene changes.

Unsurprisingly, there was a negative association between correct change detection and absence of fixations to the target items throughout the trial. Fixation of the target items, however, did not appear necessary for change detection given that in 9.7% (n= 12) and 18.5% (n=20) of detected trials, in Experiment 2a and 2b respectively, participants had correctly noticed the change without ever making an eye movement to the location of the target item. This finding suggests that the encoding of scene items and subsequent change detection can occur without direct fixation of the target items with the use of peripheral vision (which is also good at motion detection) and covert visual

Table 21. Percentage of undetected and detected trials in which participants executed each pattern of eye movement, and the Phi coefficients for the degree of association between the possible scan paths and correct change detection.

Experiment 2a	*Undetected*	*Detected*	*Phi Coefficient*
Never	66.4%	9.7%	- 0.583 * $p< .001$
Before	6.3%	5.6%	- 0.013 $p= .839$
During	7.0%	13.7%	0.110 $p= .081$
After	10.9%	29.0%	0.227 * $p< .001$
Before + During	3.1%	3.2%	0.003 $p= .964$
Before + After	1.6%	4.0%	0.075 $p= .233$
During + After	2.3%	25.8%	0.339 * $p< .001$
Before + During + After	2.3%	8.9%	0.142 * $p= .024$

Experiment 2b	*Undetected*	*Detected*	*Phi Coefficient*
Never	66.7%	18.5%	- 0.478 * $p< .001$
Before	9.0%	15.7%	0.103 $p= .103$
During	7.6%	13.0%	0.088 $p= .162$
After	8.3%	13.0%	0.075 $p= .232$
Before + During	1.4%	6.5%	0.136 * $p= .031$
Before + After	2.1%	5.6%	0.093 $p= .142$
During + After	4.9%	26.9%	0.311 * $p< .001$
Before + During + After	2.1%	9.3%	0.161 * $p= .011$

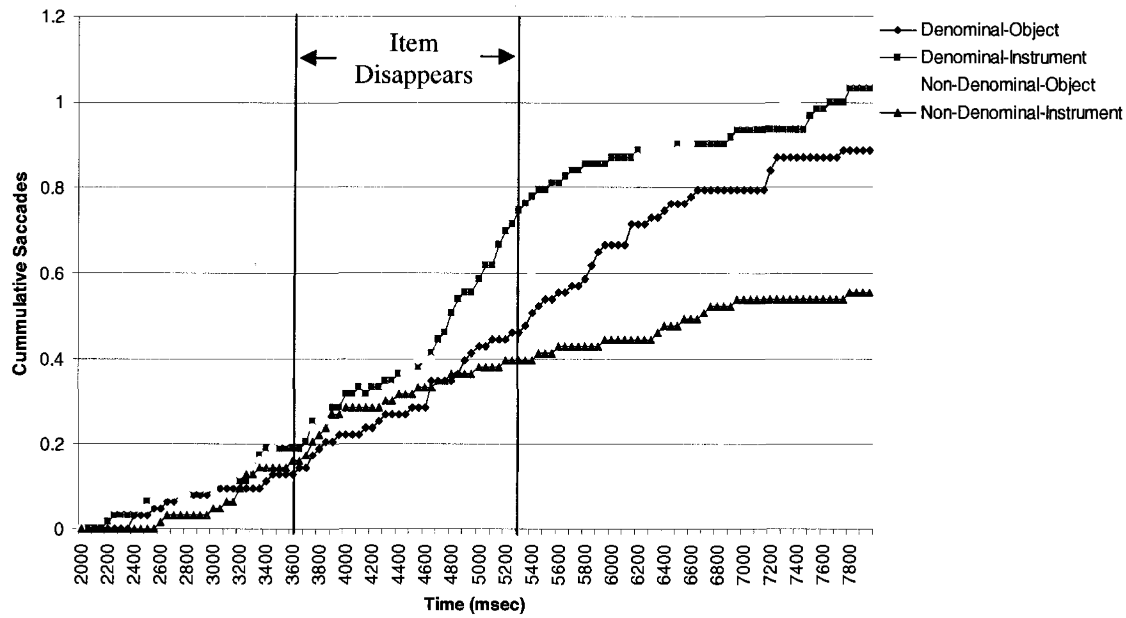
attention – particularly for Experiment 2b, in which the sentential context aided participants in correctly identifying and labeling the items in the visual scene. This assumption may also explain why successful change detection was found to occur even when the location of the target items was fixated only after the item had completely disappeared.

Moreover, in 10.9% (n= 14) and 8.3% (n= 12) of trials, for experiment 2a and 2b respectively, a scene change was not explicitly detected yet participants fixated the location of the item after it had dissolved from the screen. Since it is very rare that observers will fixate empty scene areas (Henderson & Ferreira, 2004), it is possible that the explicit detection measures underestimate the representation of change in the visual system. In these trials, either participants were cautious in reporting scene changes when in doubt of the possibility (i.e., before first detection) or the nature of the change, or the eye movements were executed without conscious awareness and/or a corresponding visual experience (i.e., the mindsight mechanism proposed by Rensink, 2004). This finding has also been reported by Hollingworth et al. (2001) who concluded that these "may reflect instances whereby participants had sufficient information to direct their eyes to the location of the change, but not enough to explicitly report the change."

2.4.2.12 Cumulative Saccades to the Target Items

Similar to the first series of experiments, for each verb type and item, the cumulative probability across trials of making saccades to the target item was calculated for each 50-ms interval from the onset to the offset of the movie. These results are plotted in Figure 57 for both studies, and appear to correspond with the pattern of correct change detection rates reported earlier. In Experiment 2a, for the items belonging to the

A)



B)

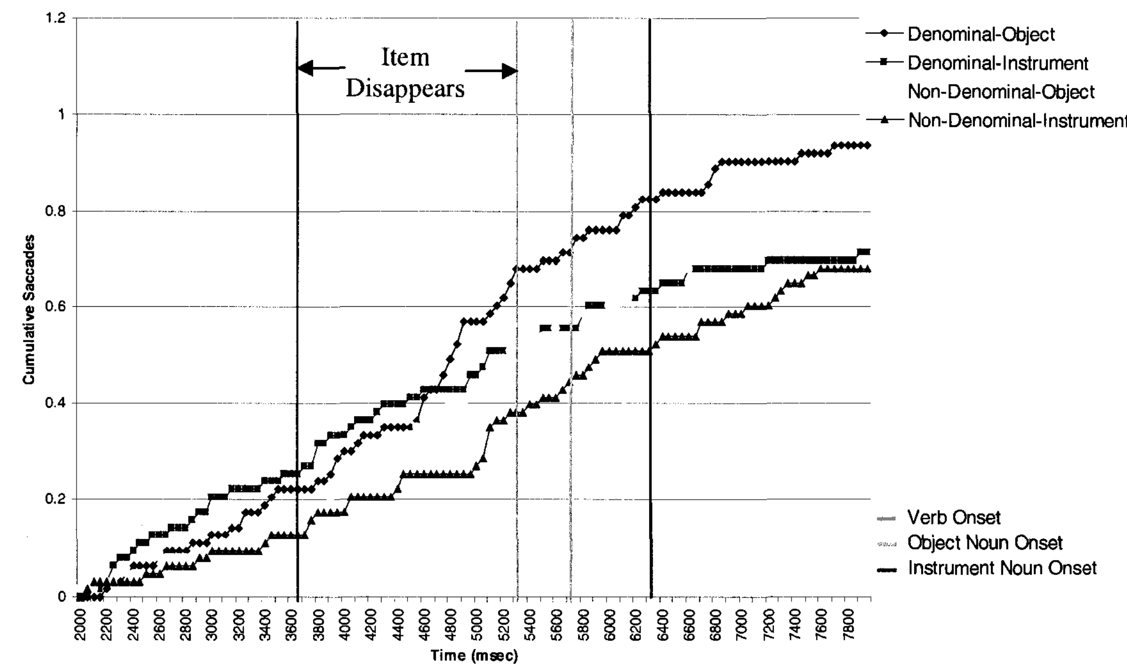


Figure 57. The cumulative saccades to the target item as a function of verb types and item types for A) Experiment 2a and B) and Experiment 2b. The disappearance onset and offset are shown, along with the verb, object and instrument noun onsets, and are aligned to the 100ms bin within which they fall.

denominal object, denominal instrument and non-denominal object categories, participants appear to have had their eyes drawn to the target items as they slowly dissolved. Moreover, saccades continued to rise after their complete disappearance, to a mean of approximately 1.0 by movie offset. By contrast, for the target items in the non-denominal instrument conditions the pattern of eye movements was initially the same, but then instead of continually rising like the other three categories, it quickly leveled off – reaching a mean of 0.55 by movie offset. It appears, therefore, that correct change detection generally occurred when participants eyes were initially drawn to the dissolving object and then maintained in the location of the item until it completely disappeared.

For Experiment 2b, for each category of target items participants appear to have had their eyes drawn to the items as they slowly dissolved. Furthermore, saccades continued to rise after the items complete disappearance, to a mean of approximately 0.75 by movie offset. The items in the non-denominal instrument category again produced lower cumulative saccade means compared to the other categories, however, contrary to the pattern of eye movements in the first experiment, cumulative saccades continued to rise until movie offset. It appears that the accompanying sentence may have guided participants to maintain their fixation on the target item until it completely disappeared or to return to its location after its complete disappearance – which in turn slightly increased change detection performance for the non-denominal instrument items as compared to Experiment 2a.

It is interesting to note that in Experiment 2b the cumulative saccades increased drastically for the denominal object items during its gradual change. These results suggest that information contained within the accompanying sentence, prior to verb onset, directed visual attention towards these items, yet did not result in overall greater change detection rates. One possibility is that the linguistic information prior to verb onset (e.g., "Later, after pouring the *flour* the man will …") may have been closely associated with the verb's complement object (e.g., eggs) or instrument (e.g., whisk) nouns. Finally, these results, along with the findings in the previous sections, appear to demonstrate that the sentential context increased looks to the less salient or more ambiguous target items (i.e., denominal objects and non-denominal instruments), while decreasing looks to the more salient and less ambiguous target items (i.e., denominal instruments and non-denominal objects).

2.4.2.13 First Post-Verb Saccades to the Target Items

Finally the average first saccade onset times to the object and instrument items relative to verb onset, in the neutral condition trials were examined, and plotted in Figure 58. A trial was excluded from the analysis if a participant was already fixating the item at verb onset. The analysis revealed a statistically significant main effect of target item, for the items analysis only, $\underline{F1}(1,7)= 3.95$, $\underline{p}= .087$, $\eta_p^2 = .361$; $\underline{F2}(1,15)= 6.06$, $\underline{p}= .026$, $\eta_p^2 = .288$, no main effect of verb type, $\underline{F1}(1,7)= 0.381$, $\underline{p}= .557$, $\eta_p^2 = .052$; $\underline{F2}(1,15)= 0.193$, $\underline{p}= .666$, $\eta_p^2 = .013$, and no significant interaction between these two factors $\underline{F1}(1,7)= 0.984$, $\underline{p}= .354$, $\eta_p^2 = .123$; $\underline{F2}(1,15)= 0.006$, $\underline{p}= .940$, $\eta_p^2 = .000$. Participants were found to be slower to initiate saccades to the target objects, as compared to their associated target instruments. These values, however, do not take into consideration that the denominal target instruments (e.g., iron), which correspond to the denominal verbs (e.g., to iron), were uttered earlier within the sentences than the non-denominal instruments (e.g., "…the woman will *iron* the shirt in the kitchen" versus "…the woman

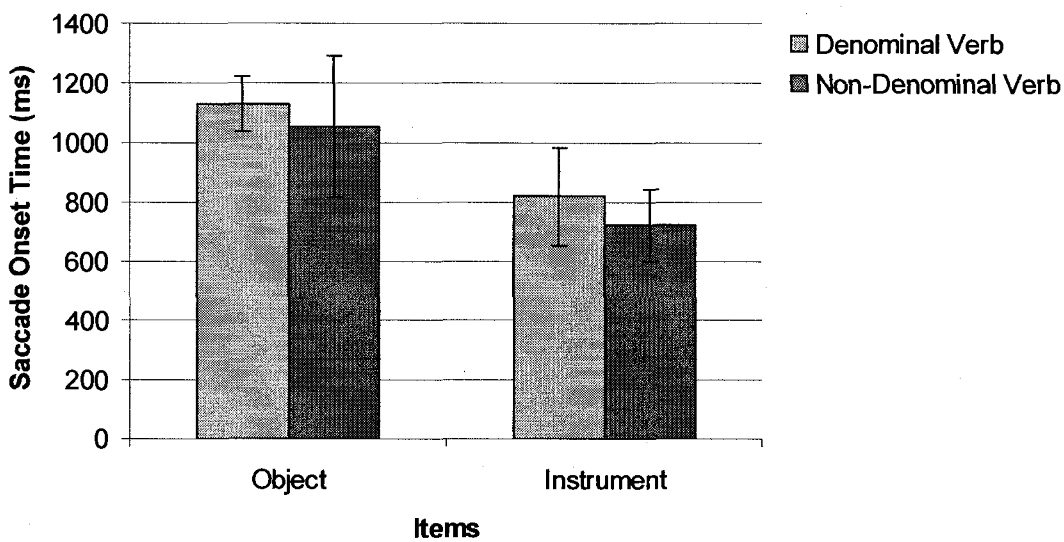


Figure 58. Average first saccade onset (SOT) to object and instrument items relative to verb onset for the no-change control condition, as a function of verb types.

will chop the vegetables with the *knife*").

Thus, to control for the significant difference in instrument noun onset across the two sentence structures, first SOTs to each target item were subtracted from their respective verbal onsets within the sentences. The differences between these two values were plotted in Figure 59. Participants in the denominal verb conditions launched saccades to the target objects and instruments approximately 739 and 820 ms after their respective onsets within the sentences. As shown in Table 22, the average denominal object and instrument noun lengths were 394 and 292 ms, suggesting that participants actually made few anticipatory eye movements to the target items – assuming it takes approximately 200 ms to program a saccade (Matin et al., 1993). This was also the case for participants in the non-denominal verb conditions, who launched their first saccades to the target nouns 663 ms after their onset within the sentences. Since the average non-denominal target nouns were 359 ms in length, these findings also suggest that participants produced few anticipatory eye movements to the target nouns via information extracted at the verb. A different story, however, emerged for the non-denominal instrument items (e.g., "…the woman will chop the vegetables with the *knife*…"). Participants in some of these trials launched their saccades 316 ms *before* the instrument noun was uttered within the sentence. As indicated in Table 22, since the average instrument noun length was 486 ms, these results would suggest that participants programmed their saccades sometime within the presentation of the object nouns.

These latter findings, however, do not appear to be a reflection of inherent properties of this set of items, as was reported above. Support for this assumption is

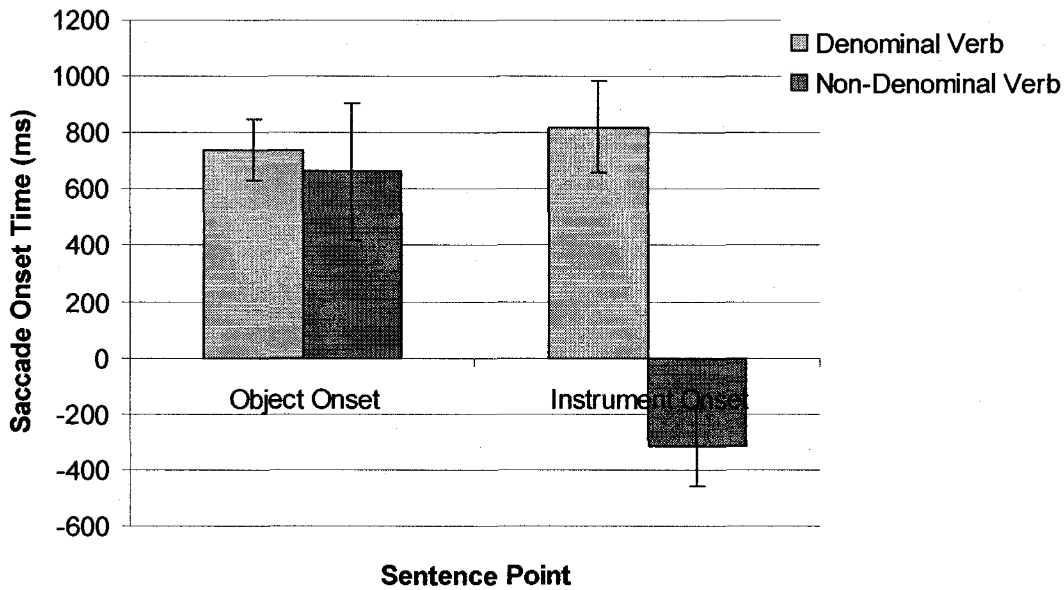


Figure 59. Average first saccade onset (SOT) to object and instrument items relative to target item onset for the no-change control condition, across verb types.

Table 22. Word durations for the denominal and non-denominal sentences (ms)

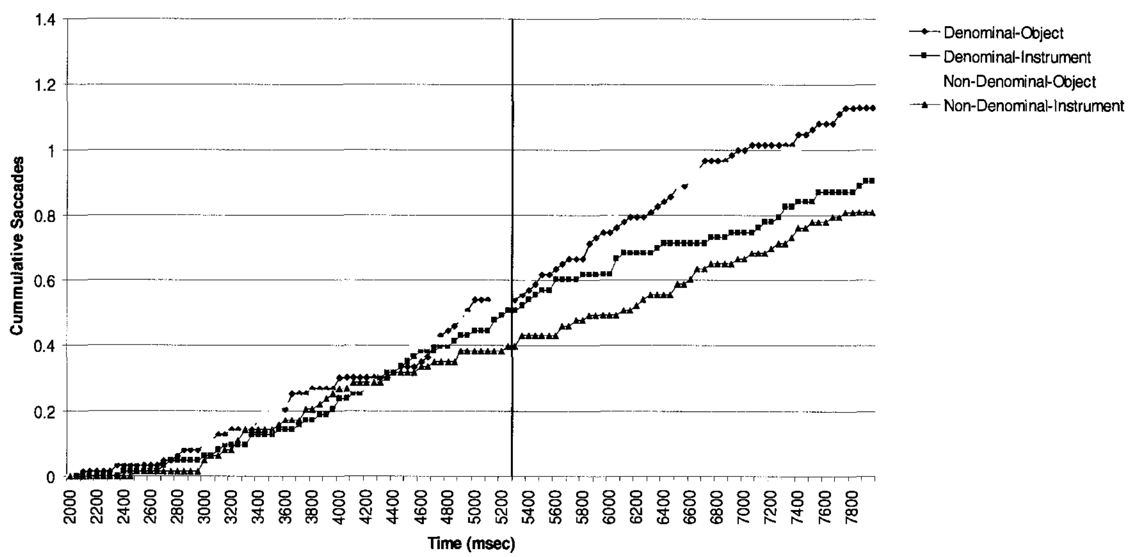
Duration	Denominal	Non-Denominal
Verb	292	245
Post-verbal break + Determiner	115	145
Target Noun	394	359
Post-noun break + Determiner	-	237
Target Instrument	-	486

further demonstrated in Figure 60, which illustrates the mean cumulative saccades to the target items in the no-change control conditions. Average cumulative saccades for the non-denominal instrument target items significantly increased after verb onset but only in Experiment 2b. Together, these results suggest that participants, in some trials, utilized the selectional restrictions of the non-denominal verbs (e.g., instrument that can chop) and/or the verb's direct object (e.g., instrument which can be used to chop vegetables), as well as the scene gist and category, to anticipate the identity and location of the explicit instruments within the scenes that will be used by the agents to perform the actions, before they are encountered within the sentence. This is what likely also resulted in the slightly higher change detection rates for these target items. It remains unclear, however, whether the information conveyed by the sentence was really used to anticipate the instrument referent or to simply confirm an event that had already been anticipated based on information extracted from the scene.

2.4.3 Discussion

In summary, overall change detection was quite poor (approximately 50%) even though the items were, for the most part, spontaneously mentioned by the participants in their verbal descriptions (~50%), and correctly identified (~90%) and recalled (~77%) in post-experimental tasks. The higher recognition rates, as compared to the change detection rates, further support the notion that the change detection task, at times, underestimated the amount of information that was contained within the viewer's visual representations. This also becomes evident when looking at the eye-movement data, which demonstrated that in a small sample of trials participants made saccades to the previous location of the target items, yet did not report noticing any change. Since it is

A)



B)

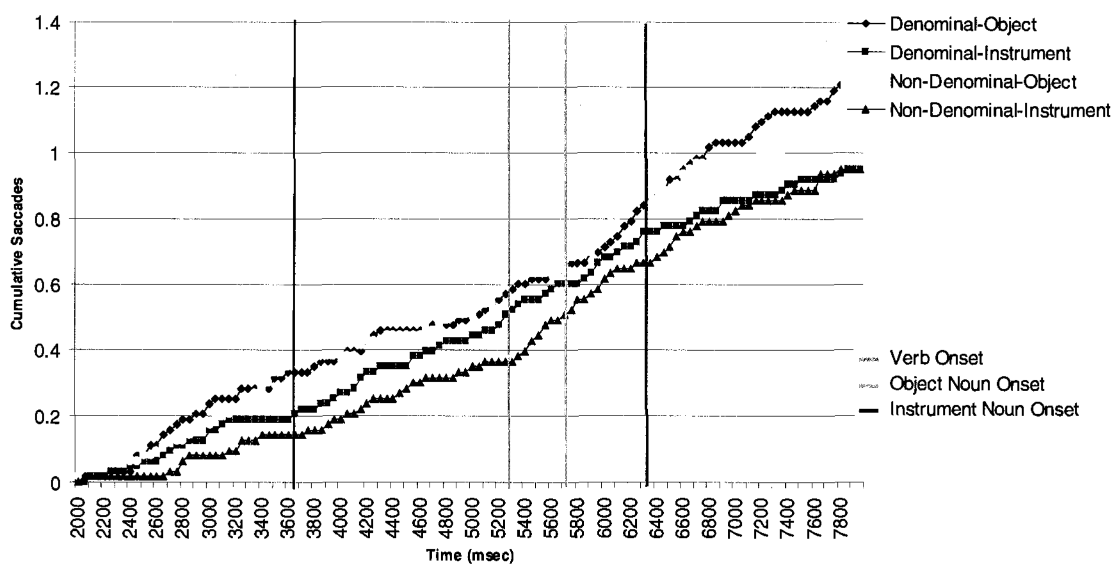


Figure 60. The cumulative saccades to the target item as a function of verb types and target items in the no-change control condition, for A) Experiment 2a and B) Experiment 2b. The verb, object and instrument noun onsets are shown, and are aligned to the 100ms bin within which they fall.

quite uncommon for viewers to make saccades to empty scene regions, this would seem to indicate that participants either had some memory trace for the changed item but were either being cautious in their response (as most occurred before the first change detection), or had insufficient information to identify the change.

Eye-movement data indicated that the poor change detection performance was in general the result of: (1) attention being drawn towards the agent in each scene (~55% of the trials) and, therefore, away from the sometimes changing target items (~15% of the trials); (2) a failure to encode the target items before they completely disappeared from the screen; and (3) a failure to either maintain long enough fixation on the target items as they were gradually dissolving (or until complete disappearance) to notice the subtle change signals, or to re-fixate the location of the target items after they had completely disappeared. Once participants became aware that items in the scene could disappear, their viewing behavior significantly changed, with higher change detection rates, greater mention and recognition of the target items, as well as increased numbers of saccades and time spent fixating the target items. These findings pose important implications with regards to change blindness studies which have used explicit change detection tasks to make inferences about the nature of our visual representations. In these studies participants may be intentionally attending and attempting to retain visual details that would normally go unnoticed. This task specific viewing behavior would then result in change detection and recognition scores that potentially overestimate the amount of visual detail we actually retain when naturally observing a scene. In other words, unless visual attention is captured by changing items or directed by the knowledge that changes can occur, observers under normal viewing circumstances may only encode the gist of the scenes, along with a few of the key items that helped determine the nature of the scenes.

Although the concurrent presentation of sentences, which described the visual scenes, significantly increased the probability of mentioning the target items, as well as correctly recalling and identifying them, they did not result in higher change detection rates or greater looks to the target items. In fact the change detection rates were slightly lower, and this did not differ by verb type. One possible explanation is that the task in the second experiment was more attentionally taxing to participants, as they were now required to not only view and memorize the visual scenes in order to later provide verbal descriptions of them, but also to process the accompanying sentences. Greater mention of the target items when describing the scene may have, for some trials, been the result of participants repeating the sentences verbatim – even though they were specifically told they need not memorize the sentences. The fact that this approach was not accompanied by greater change detection or saccades to the target items, suggests that some participants repeated what was said in the sentence without having actually perceived them in the visual scenes.

Furthermore, the higher recognition scores for the experimental trials, yet not for the no-change control trials, may indicate that either: (1) the sentential context helped participants quickly identify and label the current or previously attended target items, some of which were initially ambiguous; and/or (2) participants consciously or unconsciously answered "yes" in the recognition test, without having actually seen the target items in the visual scenes, because they either remembered them being uttered in the accompanying sentences or having repeated them during their verbal description of the scenes. Additional research using a more comprehensive recognition task, such as a two-alternative forced-choice task, may help elucidate the factor(s) responsible for this pattern of results.

Contrary to expectations the descriptive sentences did not result in more saccades or time spent fixating the location of the target items *after* they had disappeared and, thus, result in better change detection. Except for the non-denominal instrument target items, participants actually made fewer eye movements to the location previously occupied by the item, than when no sentence was given. The non-denominal instrument target items generally proved to be less easily detected, mentioned, identified and recalled by participants, possibly due to their smaller surface size (i.e., generally slim and elongated), which likely produced more subtle change signals. The sentence appears to have diminished these effects by providing labels to these ambiguous target items and subsequently directing visual attention towards the previous location of these items. In addition, the eye movement patterns did not appear to demonstrate a close association between saccade onsets and sentence onsets (i.e., time at which the item was uttered in the sentence), as saccades, except for the non-denominal instrument items, were generally initiated well after they were uttered in the sentences. These results, along with overall good recall for the target items, not only replicate the findings by Di Nardo (2005); De Almeida et al. (in preparation), but also suggest that the long SOT values were not the result of participants failing to accurately encode the target items before the onset of the verb.

Overall, these findings do not seem to be the result of participants having ignored the sentences, since their verbal responses, eye movement data, and higher identification scores indicated that they had processed the contents of the sentences. Rather, previous successful encoding of the target items, along with additional confirmation from the accompanying sentence, appear to have decreased participants' need to quickly fixate the target items as they heard the verbs and the names of the objects and instruments uttered in the sentence (e.g., "I saw the eggs mentioned in the sentence, they were on the counter"). On the other hand, with ambiguous or less salient items, linguistic variables appear to have played a greater role in guiding eye movements and further examination and confirmation tended to be required (e.g., The sentence mentioned that "…the woman will chop the vegetables with…", was that a knife I saw on the cutting board?). It remains unclear, however, whether the information conveyed by the sentence was used to anticipate the instrument referent or confirm an event that had already been anticipated based on information extracted from the scene. Furthermore, the greater frequency of change detection without prior fixation of the target items in Experiment 2b, would suggest that not only could observers encode the items without prior fixation, but also that the sentential context may have allowed them to direct covert attention and utilize peripheral vision more efficiently.

General Discussion

The purpose of this series of studies was to examine the role of visual representations in the access to arguments and conceptual structures of verbs during sentence processing. Using an eye-tracker and a change blindness paradigm, in which participants heard sentences while looking at pictures of objects (Experiment 1a,b, and c) or dynamic scenes (Experiment 2a and b), attempts were made to better understand how the language and visual systems interact to help perceive the world around us.

In the first series of studies, the purpose of the four normative studies was to ensure: (1) the appropriate selection of the preferred direct object noun for a set of lexical causative verbs; (2) that the selected objects were recognized equally well across categories (i.e., noun, phonological, semantic, color, shape and unrelated); (3) that the names assumed to be the main labels for each object were correct; and (4) that competitor items assumed to share similar features with the direct object noun were correct (i.e., sound onset, function, color and form). The purpose of Experiments 1a, 1b and 1c was to replicate Altmann and Kamide's (1999) study, which examined the timing of eye movements to objects in still pictures during spoken sentence processing, but with the use of a new eye-tracker paradigm. Two sets of sentences were contrasted, a highly constraining causative construction in which there was a close conceptual relation between the verb and its direct object, and a neutral construction with a transitive perception verb. Eye-movement data were collected at six probe points, three at the noun direct object of the main verb (Experiment 1a) and three at the main verb, with 6 (Experiment 1b) or 24 seconds (Experiment 1c) to encode the visual display.

In the second series of studies, the purpose of the two normative studies and the pilot study was to ensure: (1) the appropriate selection of the preferred direct object noun and instrument for a list of denominal and non-denominal verbs; (2) that the gradual change paradigm produced similar change detection rates as the commonly used flicker paradigm; and (3) that gradual changes were equally noticeable for the target object and instrument items in each scene. The purpose of Experiments 2a and 2b was to replicate the findings by de Almeida, et al. (in preparation), which examined the timing of eye movements to objects in dynamic scenes during spoken sentence processing, but with the additional use of a change blindness paradigm. Eye movement timing and patterns, as well as correct change detection and recognition, were examined as participants viewed dynamic scenes, in which the grammatical complement or instrument of the verb gradually dissolved from the screen. In Experiment 2b, participants were also presented with a related sentence that contained a verb which either explicitly or implicitly named the instrument that would perform the event.

Summary of Results

In the first series of studies, when participants were probed at the noun (Experiment 1a), the more semantically restrictive causative verbs did not result in fewer first saccade errors or quicker saccade onset times to the target object, as compared to the less semantically restrictive perception verbs. In addition, participants did not make fewer first saccade errors or quicker saccade onset times to the target noun when it was embedded within a sentence structure, than when presented in isolation. By the advent of the noun, information extracted from the more semantically restrictive verb may have constrained the domain of reference in the scene, but participants seem to have delayed saccades by a few hundred milliseconds to when the noun's uniqueness point was reached and, thus, the target object was confirmed. This cautious strategy seems warranted in view of the presence of phonological and semantic competitors, who were often mistaken for the target object (9.1% and 12.5% respectively). In fact, across all three experiments, early within the processing of the verb-noun complement, objects that were related to the target noun along both phonological and semantic dimensions appear to have become activated. By contrast, during the later stages of processing, stored information about visual form and color appear to have become activated and possibly used to guide visual attention towards objects in the display that shared the same visual features as the target object.

When probed at the verb and given 6 seconds to initially scan the visual display (Experiment 1b), participants made significantly more first saccade errors than in Experiment 1c, when given 24 seconds to scan the visual display. These results suggest that when given less time to consolidate the visual representations into LTM, participants were more likely to use the semantically restrictive nature of the causative verbs to produce anticipatory eye movements to the most probable direct objects of the verbs. Saccades were launched before the onset of the noun in 27.0%, 21.6% and 9.77% of the causative trials, compared to 28.7%, 11.2% and 5.17% of the perception trials, for the verb onset, middle and offset conditions respectively. Moreover, saccades were launched towards the target object in the causative trials on average 32 ms, 78 ms and 321 ms after the onset of the noun for the verb onset, middle and offset conditions, respectively, as compared to 156 ms, 211 ms, and 269 ms in the perception trials. If we take the programming of a saccade to necessitate approximately 200 ms, then the saccades for the verb onset and middle conditions were actually initiated during the lifetime of the verb. Thus, during the utterance of the verb, visual attention seems to have been automatically guided by the semantic restrictions of the causative verb. By verb offset, however, with the syntactic knowledge that the target noun would be uttered next, it appears that participants employed a similar strategy as in Experiment 1a.

Finally in Experiment 1c, in which participants were given more time to pre-scan the visual display, there was a significant reduction in first saccade errors, yet this did not result in quicker mean SOT – in fact correct first SOTs were significantly slower, as compared to eye movement performances in Experiment 1b. In this experiment there appears to have been a speed-accuracy tradeoff, with participants producing fewer anticipatory eye movements and waiting until the target nouns were verbally encountered before initiating their first saccades. Saccades were launched towards the target objects before the onset of the noun in only 20.4%, 17.1% and 5.55% of the causative trials, compared to 17.1%, 11.1% and 2.78% of the perception trials, for the verb onset, middle and offset conditions, respectively. As in Experiment 1b, participants were quicker to initiate a saccade to the verb's direct object when it was preceded by the causative verb, but this appears to represent a confirmatory effect, rather than an anticipatory one.

In the second series of studies, overall change detection was quite poor (approximately 50%) even though the items were, for the most part, spontaneously mentioned by the participants in their verbal descriptions (~50%), and correctly identified (~90%) and recalled (~77%) in post-experimental tasks. Eye-movement data indicated that the poor change detection performance was in general the result of attention being frequently drawn to the agent in each scene; a failure to encode the target items before they completely disappeared from the screen, and a failure either to maintain long enough fixation on the target items as they were gradually dissolving or to re-fixate the location of the target items after they had completely disappeared. Once participants became aware that items in the scene could disappear, their viewing behavior changed significantly, with higher change detection rates, greater mention and recognition of the target items, as well as increased number of saccades and time spent fixating the target items. Although the concurrent presentation of sentences that described the visual scenes significantly increased the probability of mentioning the target items, as well as correctly recalling and identifying them, they did not result in higher change detection rates or more looks to the target items. In fact the change detection rates were slightly lower, and this did not differ by verb type.

This was also evident in the eye movement data, which did not show an increase in the number of saccades or the time spent fixating the location of the target items *after* they had disappeared (i.e., verb onset). Except for the non-denominal instrument target items, participants actually made fewer eye movements to the location previously occupied by the items, than when no sentence was given. In addition, the eye movement patterns did not appear to demonstrate a close association between saccade onsets and sentence onsets, as saccades, except for the non-denominal instrument items, were generally initiated well after they were uttered in the sentences. These results, not only replicate de Almeida et al. findings (Di Nardo, 2005; de Almeida et al., in preparation), but also those originally observed by Cooper (1974), the pioneer of the visual world paradigm. In Cooper's (1974) study, results indicated that the mean percent of automatic fixations to the correct objects was only 37%, 55% of which were initiated prior to word termination and 40% within the first fifth of a second following word termination. The author argued that the relatively low percentage of automatic fixations was a result of participants continually alternating between three types of eye movement behavior (fixating the target object; free-scanning the picture, and fixating one point independent of the concurrently heard language).

In sum, all these findings suggest that syntactic and semantic information extracted at the verb can be used to constrain the domain of reference in the scene, and in some cases predict the referent of the grammatical complement of the verb *before* the semantic properties of the direct object become available. However, eye movements to the target objects, anticipatory or in general, do not appear to be mandatory or generalizable across all situations. Rather the rate of occurrence of anticipatory eye movements seems to vary as a function of the complexity and ambiguity of the visual input, the semantic restrictions offered by the sentential context, consolidation of the visual representations in LTM, the presence of concurrent competitor objects, as well as task demands.

Result Implications

The results reported in this series of studies appear to indicate that the central processor incrementally integrates all available knowledge, linguistic and non-linguistic, with the aim to rapidly interpret the linguistic description of what we see in the world and how we may interact with it. Based on these findings, a model is proposed and illustrated in Figure 61 to explain the possible sequence of processing events that occurs for the visual and auditory systems during scene viewing and spoken sentence comprehension. The assumption is that the language and visual systems process their relative input separately. The acoustic input is decoded by the linguistic system, which engages in low-level processing (i.e., phonological identification, lexical access, word recognition, and syntactic parsing); while the visual input, obtained via scene fixations, peripheral vision, and covert visual attention, is initially processed by the visual system (e.g., color, form, texture, motion, and object recognition). However, unlike acoustic stimuli, for which there is only one chance to capture the input, visual stimuli can, for the most part, be re-fixated for further analysis.

The processes of speech and visual perception are automatic; when you hear your

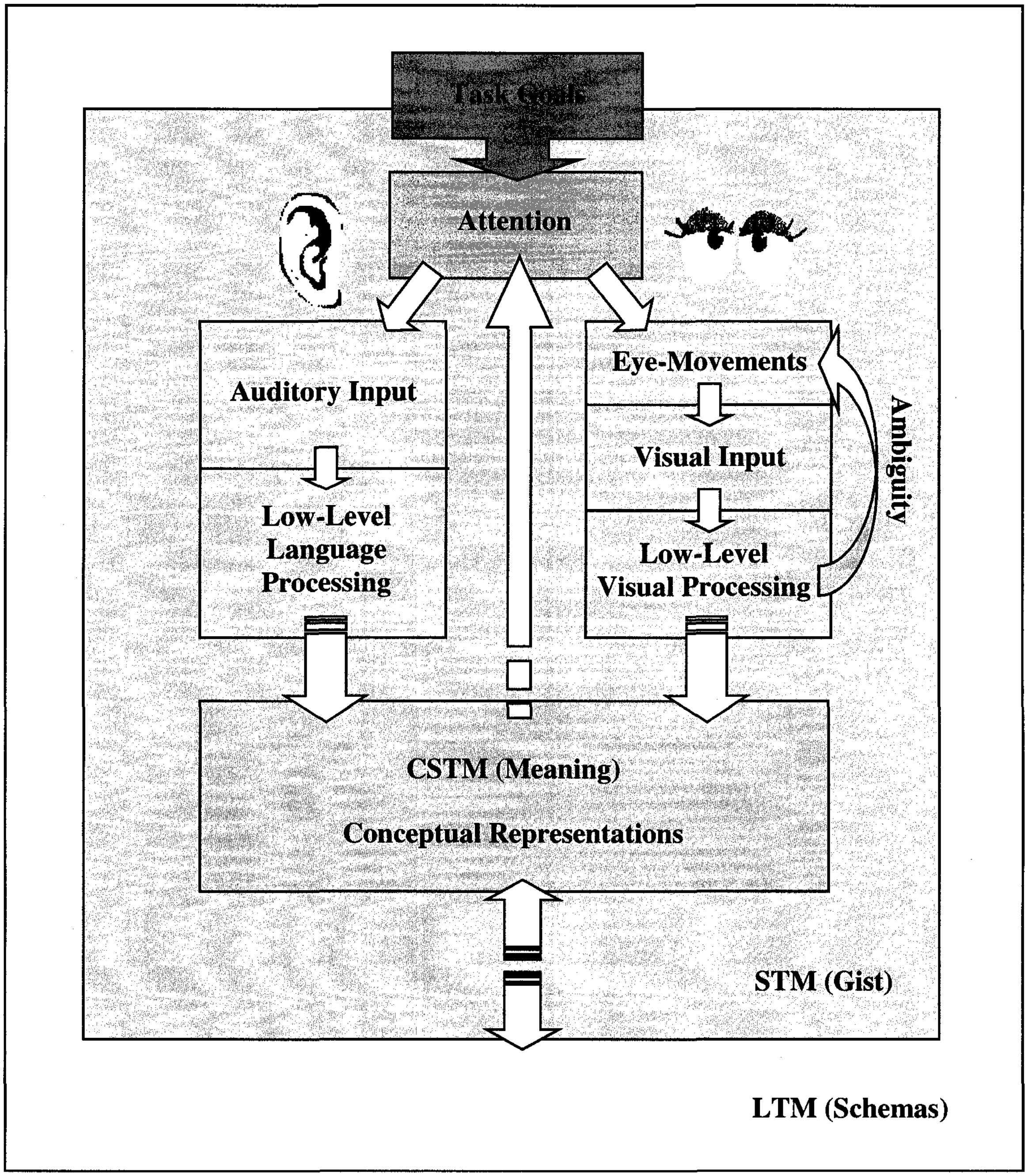


Figure 61. A model demonstrating the sequence of processing events by the visual and auditory systems during scene viewing and spoken sentence comprehension. CSTM = Conceptual Short-Term Memory; STM = Short-Term Memory; and LTM = Long-Term Memory.

name or observe a car quickly approaching, you cannot make yourself not understand or see it. In these situations attention is captured by the stimulus and the eyes and/or head are directed towards the source of the sound or movement for further processing. Yet, there are times when the attentional resources are fully allocated to another task (e.g., reading a book or talking on a cell phone), so that these unexpected stimuli do not capture attention and, thus, are not processed by the auditory and visual systems. Therefore, attention, whether stimulus-driven or directed, is necessary for the processing of perceptual input, and is contingent on task goals (see Folk, Remington, & Johnston, 1992; Folk & Remington, 1996, 2006). When engaging in a specific task, information contained within STM and LTM quickly becomes activated and plays an important part in directing attention, perceptual and response behavior, so that the end goal is effectively achieved.

Once the auditory and visual stimuli are identified by their respective processor, their meaning and conceptual representations become activated within CSTM (see Potter, 1976, 1993, 1999). CSTM then attempts to conceptually structure and integrate the activated representations by linking them with the semantically related information contained within STM (e.g., gist of the conversation or visual scene) and LTM (e.g., schemas of how to bake a cake or the appropriate location of an object), and thereby activating this information as well. When processing dynamic scenes and unfolding speech, a common occurrence in our every day lives, the information contained within the CSTM rapidly changes, such that only conceptually relevant and organized representations are selected for further processing and retained within STM and LTM. The activated information within these systems can then be used to guide what and how additional input will be processed or re-analyzed. For instance, conceptual representations activated within CSTM can be used to select alternative sentence parsings or interpretations, or what region of the scene should be inspected next in order to complete the task at hand.

Providing a few examples, based on the studies reported above, may help to further elucidate this model. In the first series of studies, participants' task was to look and indicate as quickly as possible whether the noun uttered in the accompanying sentence was present in the visual display. As the display was made up of an array of objects with no discernable gist, participants developed the task-specific strategy of fixating and memorizing each object. Since no apparent conceptual relationship existed between the objects in the visual display the activated information within CSTM made linking, structuring and consolidation of visual representations within LTM difficult and more prone to interference from subsequent conceptual information – in particular when less time was available. When faced with relatively poor and non-complex visual input, in terms of content and layout, participants appear to rely more heavily on sentence variables, in particular on specific information extracted from the verb (and in some contexts the agent), to either anticipate, accelerate or confirm the identity of the object in the visual scene that is consistent with the selectional restrictions of the verb and to make the required response (in this case a saccade and button press). The timing of this response, in relation to noun onset, varies as a function of the constraining nature of the verb, stimulus duration, competition from cohort activation, and sentence position at time of probing.

A different story, however, emerged when processing true dynamic scenes, such as was the case in the second series of studies. Here the task was to memorize the events and details of scenes in order to later describe them. The goal of this task, therefore, resulted in most of the attentional resources being allocated to processing the visual stimuli, especially after participants became aware that items within the scene could gradually dissolve. In these circumstances eye movements were, for the most part, directed or captured by the visual stimuli, in particular the moving objects or agents. This finding had already been reported long ago by Yarbus (1967), who concluded that human features, such as the eyes and lips, tend to attract viewers attention, because they tend to be mobile (bottom-up process) and rich in information (top-down process). The same, however, could be said for other human features such as the agent's hands or overall direction of motion.

In complex dynamic scenes, as opposed to arrays of objects, the conceptual information activated by the individual objects and scene gist, with the subsequent retrieval of scene schemas from LTM, can be integrated to form structured representations that are then consolidated into LTM. Participants in these circumstances appear to rely on sentence context to complement and/or confirm information that has already been extracted from the scene, and to more accurately label, describe and retain this visual information. Only when faced with poor or ambiguous visual input, do language variables appear to play a more significant role in guiding visual attention and the associated eye movements.

Limitations and Future Work

Although the new eye-tracker methodology used in the first series of experiments allowed for the control of many important variables that had not been considered in

Altmann and Kamide's (1999) study, this in turn resulted in a new set of limitations. As mentioned earlier, the use of arrays of objects or "Ersatz scenes", differ significantly from "true scenes" in how they are processed visually and cognitively. First, in true scenes there is a coherent meaningful relationship between the objects in the scene that allows for rapid identification of a scene category and subsequent retrieval of additional semantic information associated with that category. This is not possible with an array of objects that contains no natural structure and, therefore, no higher level meaning beyond that provided by the individual objects. Second, objects in arrays tend to violate proportionality properties and spatial licensing constraints. Moreover, the absence of background detail in arrays often results in larger objects, as the scenes are less dense and objects are not occluded.

Another concern was that the instructions given to participants in this series of studies may have resulted in eye movement behavior that was not representative of actual every day language processing. The task demand, to respond as quickly as possible to whether the noun object referred to in the sentence was present in the visual display, may have prompted participants to "artificially" anticipate the identity of the target noun before it was uttered in the sentence. In addition, asking participants to hold back their eye movements until the onset of the green dot, as well as perform a judgment task, placed additional attentional and processing demands, which may have interfered with language processing and, therefore, question the ecological validity of these studies. All these concerns were, to some extent, resolved in the second series of studies with the use of true dynamic scenes. When using true dynamic scenes, however, an unavoidable tradeoff occurs between ecological validity and experimental control, as was the case in the second series of experiments.

Although measures were taken to ensure that the best stimuli were used, by conducting two normative studies and one pilot study, the items selected to represent the non-denominal instruments, by chance alone, were found to be less identifiable and visually salient than the items in the other three categories. Rendering all items in a true scene equally salient (e.g., in size, color, luminance, etc.), however, is a difficult task, in particular when they were originally selected to meet certain linguistic criteria. For this reason it is imperative to run one or more control studies to obtain baseline measures against which the experimental values can be compared in order to evaluate the effects of the independent variable(s) – thus, controlling for any confounding variables. An additional control study, in which all movies are presented unedited (i.e., without scene changes) along with their related sentences, would also prove helpful. While similar to the no-change control condition trials, the values obtained from this experiment would not suffer from cross trial contamination as a result of eventual knowledge that items in the scene could gradually disappear. The obtained baseline values from this study could then be contrasted against the values obtained in Experiment 2b to determine the main effect of scene changes on sentence comprehension and dynamic scene viewing behavior.

It could be argued that it is not entirely clear whether the effects found in these studies reflect true effects of context on the processing of linguistic-specific information or whether they are task artifacts. As raised by Tanenhaus, Magnuson, Dahan, and Chambers (2000), "…the most serious concern is that the combination of a circumscribed visual world and restricted set of instructions encourages participants to develop task-specific strategies that bypass "normal language processing." Task-specific strategies appear to have been employed across the two series of studies, with associated viewing behavior varying as a function of visual and linguistic variables, such as complexity, ambiguity and duration. For instance, in the second series of studies, some participants may have eventually realized that the items referred to in the accompanying sentences were always found to the left and right side of the agent. Thus, location of the items, rather than sentence or visual context, would have been used to anticipate the object referents. The results, therefore, need to be interpreted with care, especially when attempting to generalize the results to real-world environments where no such artificial constraints and predictability exist. The use of more naturalistic stimuli and tasks, such as dynamic scenes, is an important step, however, towards simulating conditions that will elicit and allow for the study of normal language processing in a laboratory setting.

Finally, the finding that visual attention can be shifted without a corresponding shift in fixation location (Posner, 1980; Fischer & Breitmeyer, 1987) poses an important limitation on the visual world paradigm. In the two series of experiments, participants' visual attention at verb onset may have shifted in the direction of the target object, but without an associated eye movement. Thus, with the use of covert local and global attention, as well as peripheral vision, participants could have extracted a large amount of information in few fixations and without foveation. Further support for this assumption could be found in the second series of experiments, in which some participants were able to detect scene changes without ever fixating the target object. It is, therefore, quite probable that in some trials the verb's selectional restrictions guided visual attention to the target object, but could not be detected because of the employed methodology. In order to reduce the effects of this limitation, Di Nardo and colleagues are conducting similar studies whereby the visual displays are being projected onto a large screen. This mode of presentation will first of all better simulate every day scene perception, and secondly increase the visual angle, which will hopefully force participants to move their eyes when inspecting the scene, rather than only shifting their visual attention.

In closing, eye-tracking research examining the role of visual representations in the access to argument and conceptual structures of verbs during sentence processing is relatively new and certainly requires further investigation. It is thought that the new methodologies developed and tested in this series of experiments, particularly the use of true dynamic scenes, will prove especially useful in conducting this additional research. Although the present studies provided valuable information concerning the functioning of the linguistic and visual systems, it remains unclear *when* and *where* the interaction between the two cognitive systems occurs during language comprehension. Use of a more systematic approach whereby each variable is manipulated and introduced incrementally, as opposed to the multi-variable exploratory approach used in this paper, will hopefully help to resolve this debate.

References

Allopenna, P.D., Magnuson, J.S., & Tanenhaus, M.K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38,* 419-439.

Altmann, G., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition, 30,* 191-238.

Altmann, G.T.M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73,* 247-264.

Altmann, G.T.M., & Kamide, Y. (2004). Now you see it, now you don't: Mediating the mapping between language and visual world. In Henderson, J., & Ferreira, F. (eds.), *The Interface of Language, Vision, and Action: Eye Movements and the Visual World.* New York: Psychology Press, 347-386.

Angelone, B.L., Levin, D.T., & Simons, D.J. (2003). The relationship between change detection and recognition of centrally attended objects in motion pictures. *Perception, 32,* 947-962.

Antes, J. R. (1974). The time course of picture viewing. *Journal of Experimental Psychology,* 103(1), 62-70.

Ballard, D.H., Hayhoe, M.M., & Pelz, J.B. (1995). Memory representations in natural tasks. *Journal of Cognitive Neuroscience, 7(1),* 66-80.

Beck, M.R., & Levin, D.T. (2003). The role of representational volatility in recognizing pre- and postchange objects. *Perception & Psychophysics, 65(3),* 458-468.

Biederman I, (1981). On the semantics of a glance at a scene. In Kubovy, M., Pomerantz, J.R. (Eds.), *Perceptual Organization* (pp 213-253). Hillsdale, NJ: Lawrence Erlbaum Associates.

Biederman, I. (1988). Aspects and extensions of a theory of human image understanding. In Z. W. Pylyshyn (Ed.), *Computational Process in Human Vision: An Interdisciplinary Perspective* (pp 370-428). Norwood, NJ: Ablex.

Bridgeman, B., Hendry, D., & Stark, L. (1975). Failure to detect displacement of the visual world during saccadic eye movements. *Vision Research, 15(6),* 719-722.

Buswell, G.T. (1935). *How People Look at Picture.* Chicago: University of Chicago Press.

Carpenter, R.H.S. (1988). *Movements of the Eyes.* (2nd ed.). London: Pion.

Chambers, C.G., Tanenhaus, M.K., Eberhard, K.M., Filip, H., & Carlson, G.N. (2002). Circumscribing referential domains during real-time language comprehension. *Journal of Memory and Language, 47,* 30-49.

Chomsky, N. (1965). *Aspects of the theory of syntax.* Oxford, England: M.I.T. Press.

Clark, E.V., & Clark, H.H. (1979). When Nouns Surface as Verbs. *Language* 55, 767-811.

Clark, H.H. (1992). *Arenas of language use.* Stanford, CA: Center for the Study of Language & Information.

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

Clifton, C., Frazier, L., & Connine, C. (1984). Lexical expectations in sentence comprehension. *Journal of Verbal Learning and Verbal Behaviour, 23,* 696-708.

Cooper, R.M. (1974). The control of eye fixation by the meaning of spoken language. *Cognitive Psychology, 6,* 84-107.

Dahan, D., Magnuson, J.S., Tanenhaus, M.K., & Hogan, E.L. (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes, 16(5-6),* 507-534.

Dahan, D., Magnuson, J.S., & Tanenhaus, M.K. (2001). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology, 42(4),* 317-367.

De Almeida, R., Di Nardo, J., & von Grünau, M. (in preparation). Understanding sentences in dynamic scenes: Evidence for modularity.

De Graef, P., Christiaens D., & d'Ydewalle, G. (1990). Perceptual effects of scene context on object identification. *Psychological Research, 52(4),* 317-29.

Di Nardo, J.C. (2005). Eye movements as a function of spoken sentence comprehension and scene perception. *Master's Dissertation.* Concordia University, Montreal, Quebec.

Eberhard, K.M., Spivey-Knowlton, M.J., Sedivy, J.C., & Tanenhaus, M.K. (1995). Eye movements as a window into real-time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research, 24,* 409-436.

Fairchild, M. (1998). *Color Appearance Models*. Reading, Massachusetts: Addison, Wesley, & Longman.

Ferreira F., & Clifton C. (1986). The independence of syntactic processing. *Journal of Memory & Language, 25,* 75–87.

Fischer, B., & Breitmeyer, B. (1987). Mechanisms of visual attention revealed by saccadic eye movements. *Neuropsychologia, 25,* 73-83.

Fodor, J.A. (1983). *The Modularity of Mind: An Essay on Faculty Psychology.* Cambridge, MA: The MIT Press.

Fodor, J.A. (2000). *The Mind Doesn't Work That Way: The Scope and Limits of Computational Psychology.* Cambridge MA: The MIT Press.

Folk, C.L., Remington, R.W, & Johnston, J.C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance, 18(4),* 1030-1044.

Folk, C.L., & Remington, R.W. (1996). When knowledge does not help? Limitations on the flexibility of attentional control. In Kramer, A.F., Coles, M.G.H., & Logan, G.D. (eds.), *Converging operations in the study of visual selective attention (pp. 271-295).* Washington, D.C.: American Psychological Association.

Folk, C.L., & Remington, R. (2006). Top-down modulation of preattentive processing: Testing the recovery account of contingent capture. *Visual Cognition, 14(4-8),* 445-465.

Frazier, L. (1987). Theories of sentence processing. In Garfield, J.L. (ed.), *Modularity in knowledge representation and natural-language understanding (pp. 291-307).* Cambridge, MA: The MIT Press.

Frazier, L. (1989). Against lexical generation of syntax. In Marslen-Wilson, W. (ed.), *Lexical representation and process* (pp. 505-528). Cambridge, MA: The MIT Press.

Friedman, A. (1979). Framing pictures: The role of knowledge in automatized encoding and memory for gist. *Journal of Experimental Psychology: General, 108(3),* 316-355.

Garfield, J. L. (1987). *Modularity in knowledge representation and natural language understanding.* Cambridge, MA: The MIT Press.

Grimes, J. (1996). On the failure to detect changes in scenes across saccades. In Akins, K.A. (ed.), *Perception* (pp. 89-110). New York, NY: Oxford University Press, 1996.

Grimshaw, J. (1990). *Argument structure.* Cambridge, MA: The MIT Press.

Henderson, J. M. (1996). Visual attention and the attention-action interface. In K. Aikens (Ed.), *Perception: Vancouver Studies in Cognitive Science (Vol V.).* Oxford: Oxford University Press.

Henderson, J.M., & Hollingworth, A. (1998). Eye Movements during Scene Viewing: An Overview. In G Underwood (Ed.), *Eye Guidance while Reading and While Watching Dynamic Scenes.* (pp. 269-293). Oxford: Elsevier.

Henderson, J. M., & Hollingworth, A. (1999). The role of fixation position in detecting scene changes across saccades. *Psychological Science, 5,* 438-443.Henderson,

J.M., Weeks, P.A., & Hollingworth, A. (1999). Effects of semantic consistency on eye movements during scene viewing. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 210-228.

Henderson, J. M., & Hollingworth, A. (2003). Eye movements and visual memory: Detecting changes to saccade targets in scenes. *Perception & Psychophysics, 65, 58-71.*

Henderson, J. M., & Ferreira, F. (2004). Scene perception for psycholinguists. In Henderson, J.M. and Ferreira, F. (Eds.), *The interface of language, vision, and action: Eye movements and the visual world* (pp. 1-58). New York: Psychology Press.

Hollingworth, A. (2003). Failures of retrieval and comparison constrain change detection in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance, 29(2),* 388-403.

Hollingworth, A., & Henderson, J.M. (2000). Semantic informativeness mediates the detection of changes in natural scenes. *Visual Cognition (Special Issue on Change Blindness and Visual Memory), 7,* 213-235.

Hollingworth, A., Schrock, G., & Henderson, J.M. (2001). Change detection in the flicker paradigm: The role of fixation position within the scene. *Memory & Cognition, 29, 296-304.*

Hollingworth, A., Williams, C. C., & Henderson, J. M. (2001). To see and remember: Visually specific information is retained in memory from previously attended objects in natural scenes. *Psychonomic Bulletin & Review, 8,* 761-768.

Hollingworth, A., & Henderson, J. M. (2002). Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance, 28,* 113-136.

Hollingworth, A., & Henderson, J. M. (2003). Testing a conceptual locus for the inconsistent object change detection advantage in real-world scenes. *Memory &*

Cognition, 31, 930-940.

Huettig, F., & Altmann, G.T.M. (2005). Word meaning and the control of eye fixation: Semantic competitor effects and the visual world paradigm. *Cognition, 96(1),* B23-B32.

Intraub, H. (1999). Understanding and remembering briefly glimpsed pictures: Implications for visual scanning and memory. In Coltheart, V. (ed.), *Fleeting Memories: Cognition of Brief Visual Stimuli* (pp 47–70). Cambridge, MA: The MIT Press.

Irwin, D.E. (1991). Information integration across saccadic eye movements. *Cognitive Psychology, 23(3),* 420-456.

Irwin, D.E., Yantis, S., & Jonides, J. (1983). Evidence against visual integration across saccadic eye movements. *Perception & Psychophysics, 34(1),* 49-57.

Jackendoff, R. (1990). *Semantic structures.* Cambridge, MA: The MIT Press.

Jonides, J., Irwin, D.E., & Yantis, S. (1983). Failure to integrate information from successive fixations. *Science, 222(4620),* 188.

Kahneman, D., Treisman, A., Gibbs, B.J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology, 24(2),* 175-219.

Kamide, Y., Altmann, G.T.M., & Haywood, S.L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language, 49,* 133-156.

Levin, B. (1993). Introduction: The theoretical perspective. In *English Verb Classes and Alternations: A preliminary Investigation* (pp. 1-23). Chicago: University of Chicago Press.

Levin, D.T., & Simons, D.J. (1997). Failure to detect changes to attended objects in motion pictures. *Psychonomic Bulletin and Review, 4,* 501-506.

Levin, D.T., Simons, D.J., Angelone, B., & Chabris, C.F. (2002). Memory for centrally attended changing objects in an incidental real-world change detection paradigm. *British Journal of Psychology, 93,* 289-302.

Loftus G.R., & Mackworth N.H. (1978). Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human Perception and Performance, 4(4)*, 565-72.

Loftus, E.F. (1979). Reactions to blatantly contradictory information. *Memory & Cognition, 7(5),* 368-374.

MacDonald, M.C., Pearlmutter, N.J., & Seidenberg, M.S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review, 101(4),* 676-703.

Mackworth, N.H., & Bruner, J. S. (1970). How adults and children search and recognize pictures. *Human Development,* 13, 149-177.

Mackworth, N.H., & Morandi, A.J. (1967). The gaze selects informative details within pictures. *Perception and Psychophysiology, 2,* 547-552.

Mannan, S. K., Ruddock, K. H., & Wooding, D. S. (1995). Automatic control of saccadic eye movements made in visual inspection of briefly presented 2-D images. *Spatial Vision, 9,* 363-386.

Marslen-Wilson, W. (1990). Activation, competition, and frequency in lexical access. In Altmann, G.T.M. (ed.), *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives* (pp. 148-172). Cambridge, MA: The MIT Press.

Marslen-Wilson, W. & Tyler, L.K. (1989). Against modularity. In Garfield, J.L. (ed.), *Modularity in Knowledge Representation and Natural-Language Understanding* (pp. 37-62). Cambridge, MA: The MIT Press.

Matin, E. (1974). Saccadic suppression: A review and an analysis. *Psychological Bulletin, 81*, 899-917.

Matin, E., Shao, K., & Boff, K. (1993). Saccadic overhead: information processing time with and without saccades. *Perception and Psychophysics, 53,* 372-380.

McConkie, G.W., & Currie, C.B. (1996). Visual stability across saccades while viewing complex pictures. *Journal of Experimental Psychology: Human Perception and Performance, 22(3),* 563-581.

McConkie, G.W., & Zola, D. (1979). Is visual information integrated across successive fixations in reading? *Perception & Psychophysics, 25(3),* 221-224.

Meyer, D.E., & Schvaneveldt, R.W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology, 90(2)*, 227-234.

Mitchell, D.C. (1989). Verb guidance and other lexical effects in parsing. *Language and Cognitive Processes, 4(3-4),* SI 123-154.

Mitroff, S.R., Simons, D.J., & Levin, D.T. (2004). Nothing compares two views: Change blindness can occur despite preserved access to the changed information. *Perception & Psychophysics, 66,* 1268-1281.

Oliva, A., & Schyns, P.G. (2000). Diagnostic colors mediate scene recognition. *Cognitive Psychology, 41(2),* 176-210.

O'Regan, J.K. (1992). Solving the 'real' mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology Revue Canadienne de Psychologie, 46(3),* 461-488.

O'Regan, J.K., Rensink, R.A., & Clark, J.J. (1999). Change-blindness as a result of 'mudsplashes.' *Nature, 398(6722),* 34.

O'Regan, J.K., Deubel, H., & Clark, J.J. (2000). Picture changes during blinks: Looking without seeing and seeing without looking. *Visual Cognition, 7(1-3),* 191-211.

Pashler, H. (1988). Familiarity and visual change detection. *Perception & Psychophysics, 44(4),* 369-378.

Phillips, W.A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception & Psychophysics, 16(2),* 283-290.

Pollack, I., & Spence, D. (1968). Subjective pictorial information and visual search. *Perception and Psychophysics,* 3(1B), 41-44.

Pollatsek, A., Rayner, K., & Collins, W.E. (1984). Integrating pictorial information across eye movements. *Journal of Experimental Psychology: General, 113(3),* 426-442.

Posner, M.I. (1980). Orienting of attention. *The Quarterly Journal of Experimental Psychology, 32(1),* 3-25.

Potter, M.C. (1975). Meaning in visual search. *Science, 187,* 965-966.

Potter, M.C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory, 2*, 509-522.

Potter, M.C. (1993). Very short-term conceptual memory. *Memory & Cognition, 21,* 156-161.

Potter, M.C. (1999). Understanding sentence and scenes: The role of conceptual short-term memory. In Coltheart, V. (ed.), *Fleeting Memories: Cognition of Brief Visual Stimuli* (pp. 11-46). Cambridge, MA: The MIT Press.

Rayner, K., & Pollatsek, A. (1983). Is visual information integrated across saccades? *Perception & Psychophysics, 34(1),* 39-48.

Rayner, K., & Pollatsek, A. (1992). Eye movements and scene perception. *Canadian Journal of Psychology Revue Canadienne de Psychologie, 46(3),* 342-376.

Rensink, R.A. (2000a). Seeing, sensing, and scrutinizing. *Vision Research, 40(10-12),* 1469-1487.

Rensink, R.A. (2000b). The dynamic representation of scenes. *Visual Cognition, 7(1-3),* 17-42.

Rensink, R.A. (2002). Change detection. *Annual Review of Psychology, 53(1),* 245-277.

Rensink, R.A. (2004). Visual sensing without seeing. *Psychological Science, 15(1),* 27-32.

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(5),* 368-373.

Rensink, R.A., O'Regan, J.K., & Clark, J.J. (2000). On the failure to detect changes in scenes across brief interruptions. *Visual Cognition, 7(1-3),* 127-145.

Schafer, J.L. (1997). Analysis with missing data in prevention research. In Bryant, K.J., Windle, M., West, S.G. (Eds), *The science of prevention: Methodological advances from alcohol and substance abuse research.* Washington, DC, US: American Psychological Association (pp. 325-366).

Schyns, P.G., & Oliva, A. (1994). From blobs to boundary edges: Evidence for time and spatial scale dependent scene recognition. *Psychological Science, 5,* 195-200.

Sedivy, J.C., Tanenhaus, M.K., Chambers, C.G., & Carlson, G.N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition, 71,* 109-147.

Simons, D.J. (1996). In sight, out of mind: When object representations fail. *Psychological Science, 7,* 301-305.

Simons, D.J. (2000). Current approaches to change blindness. *Visual Cognition, 7,* 1-16.

Simons, D.J., & Levin, D.T. (1997). Change blindness. *Trends in Cognitive Science, 1,* 261-267.

Simons, D.J., & Levin, D.T. (1998). Failure to detect changes to people during a real-world interaction. *Psychonomic Bulletin and Review, 5,* 644-649.

Simons, D.J., Franconeri, S.L., & Reimer, R.L. (2000). Change blindness in the absence of visual disruption. *Perception, 29,* 1143-1154.

Simons, D.J., Chabris, C.F., Schnur, T.T., & Levin, D.T. (2002). Evidence for preserved representations in change blindness. *Consciousness and Cognition, 11,* 78-97.

Spivey, M.J., Tyler, M.J., Eberhard, K.M., & Tannenhaus, M.K. (2001). Linguistically mediated visual search. *Psychological Science, 12,* 282-286.

Spivey, M.J., Tanenhaus, M.K., Eberhard, K.M., & Sedivy, J.C. (2002). Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. *Cognitive Psychology, 45,* 447-481.

Tanenhaus, M.K., Carlson, G.N., & Trueswell, J.T. (1989). The role of thematic structures in interpretation and parsing. *Language and Cognitive Processes, 4,* SI 211-234.

Tanenhaus, M.K., Spivey-Knowlton, M.J., Eberhard, K.M., & Sedivy, J.C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268,* 1632-1634.

Tanenhaus, M.K., Magnuson, J.S., Dahan, D., & Chambers, C. (2000). Eye movements and lexical access in spoken-language comprehension: Evaluating a linking hypothesis between fixations and linguistic processing. *Journal of Psycholinguistic Research, 29,* 557-580.

Tesolin, S., & De Almeida, R. (2000). Accessing arguments and conceptual fillers of verbs. Poster Presentation. *Mental Lexicon Conference*, Montreal, Quebec.

Trueswell, J.C., Tanenhaus, M.K., & Garnsey, S.M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language, 33(3),* 285-318.

Van de Velde, C. (2001). Monitoring eye movements to uncover the role of visual context in the access to verb-complement information during sentence comprehension. *Master's Dissertation.* Concordia University, Montreal, Quebec.

Varakin, D.A., & Levin, D.T. (2006). Change blindness and visual memory: Visual representations get rich and act poor. *British Journal of Psychology, 97(1),* 51-77.

Volkman F.C. (1976). Saccadic suppression: a brief review. In: Monty R.A., Senders J.W. (eds). *Eye Movements and Psychological Processes.* (pp. 73-83). New York: Wilely.

Wallis, G., & Bülthoff, H. (2000). What's scene and not seen: Influences of movement and task upon what we see. *Visual Cognition, 7(1-3),* 175-190.

Werner, S., & Thies, B. (2000). Is 'change blindness' attenuated by domain-specific expertise? An expert-novices comparison of change detection in football images. *Visual Cognition, 7(1-3),* 163-173.

Yarbus, A.L. (1967). *Eye Movements and Vision.* New York: Plenum Press.

Yee, E., & Sedivy, J.C. (2006). Eye movements to pictures reveal transient semantic activation during spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32(1),* 1-14.

Appendix A

Stimuli Used for Normative and Pilot Studies

and Experiments 1a, 1b, 1c, 2a and 2b

Appendix A1

Gender: Female ___ Male ___ Age: ____ First Language: _______________

Please indicate (by circling the number) on a scale from 1 to 5 how similar the two objects are for one of the following features: Shape, Color, Name, or Function. "1" signifying that the objects are "not at all similar" and "5" signifying that the objects are "very similar."

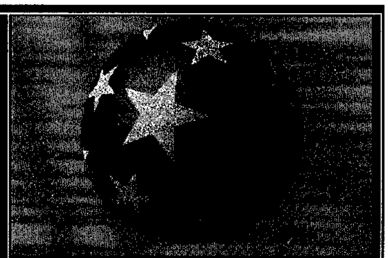
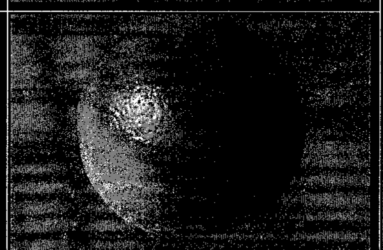
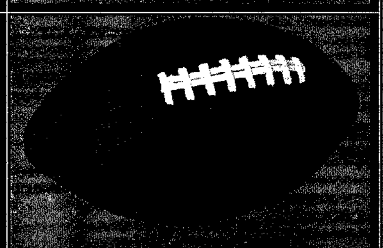
1)	1	2	3	4	5
2)	1	2	3	4	5
3)	1	2	3	4	5
4)	1	2	3	4	5
5)	1	2	3	4	5
6)	1	2	3	4	5
7)	1	2	3	4	5
8)	1	2	3	4	5
9)	1	2	3	4	5
10)	1	2	3	4	5
11)	1	2	3	4	5
12)	1	2	3	4	5
13)	1	2	3	4	5
14)	1	2	3	4	5
15)	1	2	3	4	5
16)	1	2	3	4	5
17)	1	2	3	4	5
18)	1	2	3	4	5
19)	1	2	3	4	5
20)	1	2	3	4	5
21)	1	2	3	4	5
22)	1	2	3	4	5
23)	1	2	3	4	5
24)	1	2	3	4	5
25)	1	2	3	4	5
26)	1	2	3	4	5
27)	1	2	3	4	5
28)	1	2	3	4	5
29)	1	2	3	4	5
30)	1	2	3	4	5
31)	1	2	3	4	5
32)	1	2	3	4	5
33)	1	2	3	4	5
34)	1	2	3	4	5
35)	1	2	3	4	5

--- same format up to 120

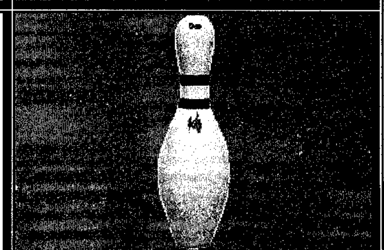
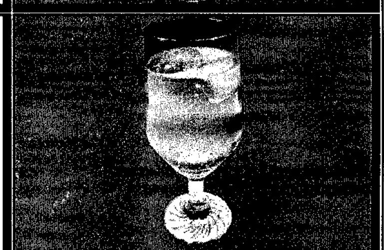
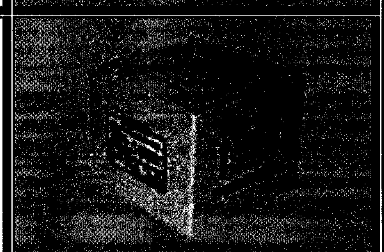
120)	1	2	3	4	5

Appendix A2

The athlete *dropped* the ball (c)
The athlete *watched* the ball (p)

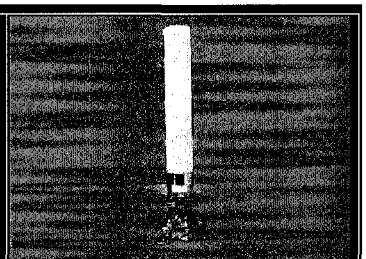
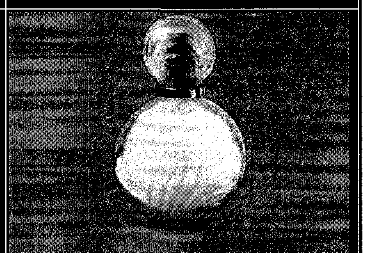
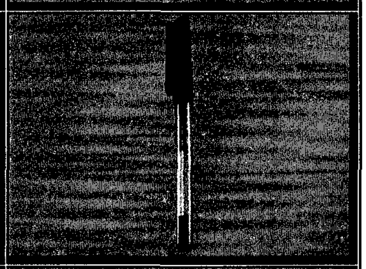
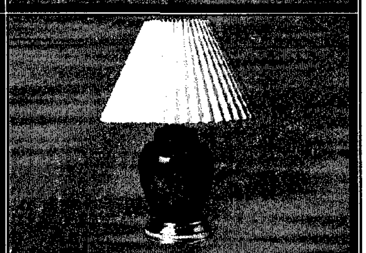
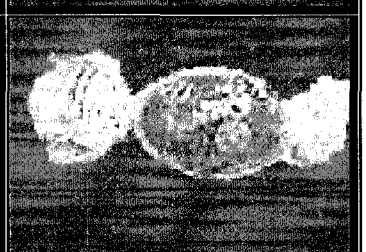
Noun	
Color	
Shape	
Semantic	
Phonological	
Unrelated	

The teenager *spun* the bottle (c)
The teenager *scented* the bottle (p)

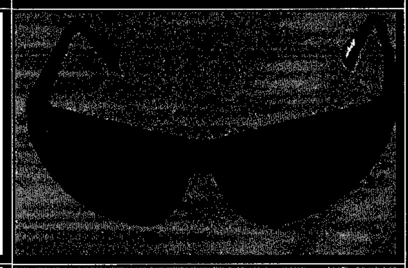
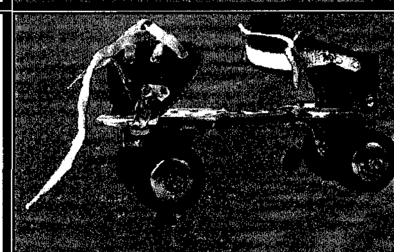
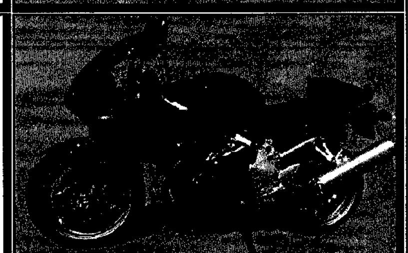
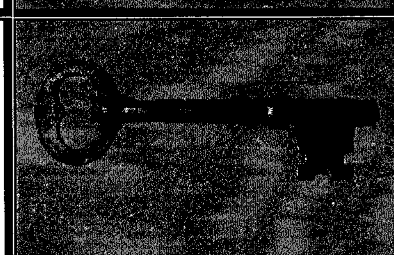
Noun	
Color	
Shape	
Semantic	
Phonological	
Unrelated	

Appendix A2

The waitress *burned* the candle (c)
The waitress *felt* the candle (p)

Noun	
Color	
Shape	
Semantic	
Phonological	
Unrelated	

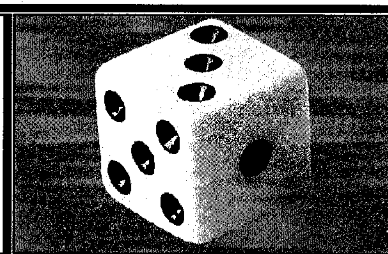
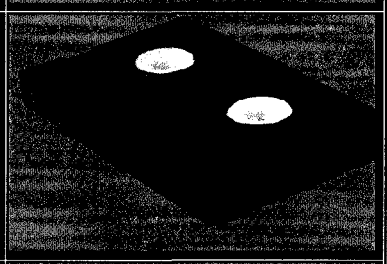
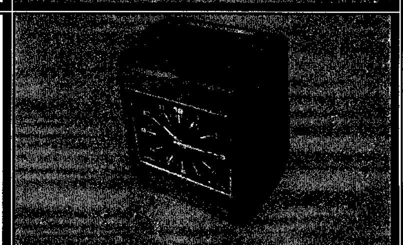
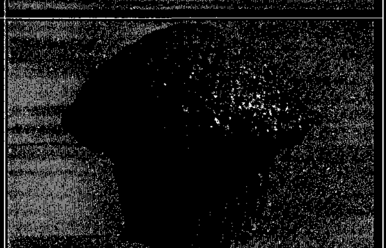
The driver *started* the car (c)
The driver *smelled* the car (p)

Noun	
Color	
Shape	
Semantic	
Phonological	
Unrelated	

Appendix A2

The baker *melted* the chocolate (c)
The baker *spotted* the chocolate (p)

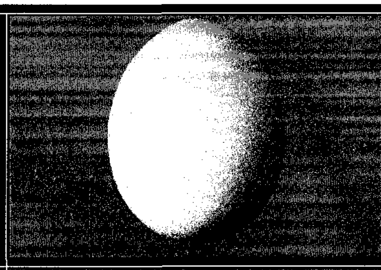
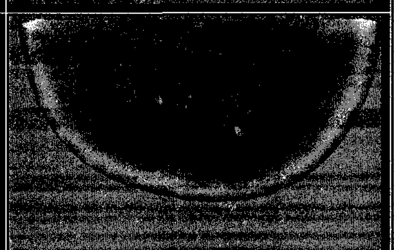
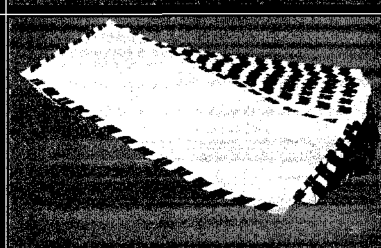
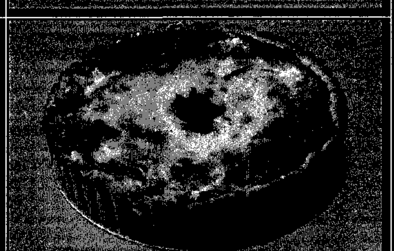
The gambler *rolled* the dice (c)
The gambler *noted* the dice (p)

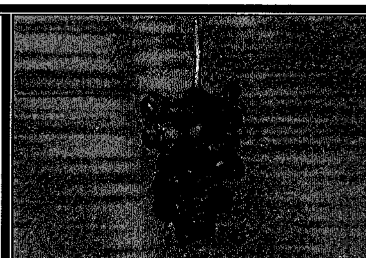
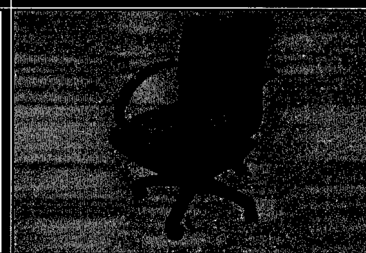
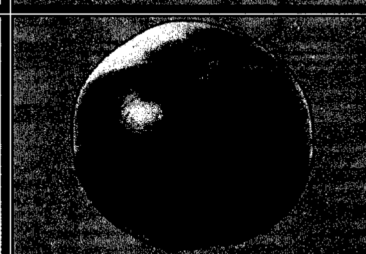
Noun		Noun	
Color		Color	
Shape		Shape	
Semantic		Semantic	
Phonological		Phonological	
Unrelated		Unrelated	

Appendix A2

The cook *cracked* the egg (c)
The cook *scanned* the egg (p)

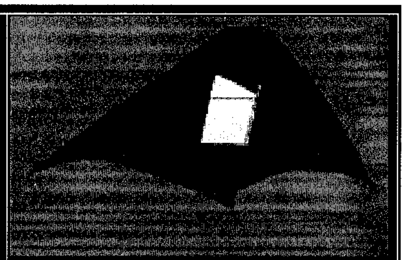
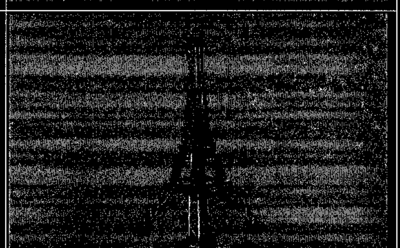
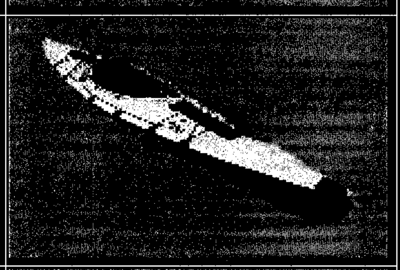
The farmer *crushed* the grapes (c)
The farmer *spied* the grapes (p)

Noun	
Color	
Shape	
Semantic	
Phonological	
Unrelated	

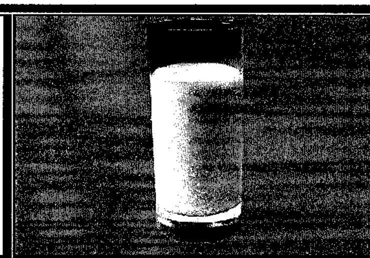
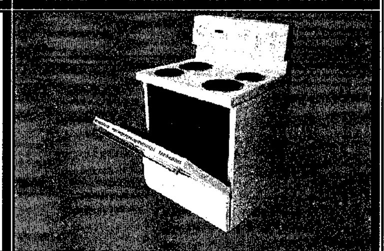
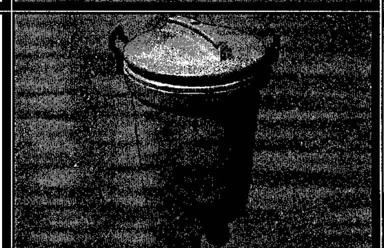
Noun	
Color	
Shape	
Semantic	
Phonological	
Unrelated	

Appendix A2

The child *flew* the kite (c)
The child *viewed* the kite (p)

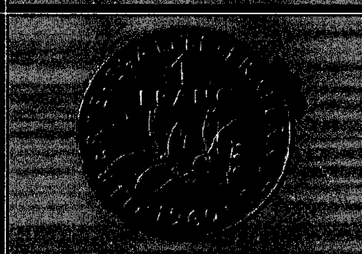
Noun	
Color	
Shape	
Semantic	
Phonological	
Unrelated	

The baby *spilled* the milk (c)
The baby *sensed* the milk (p)

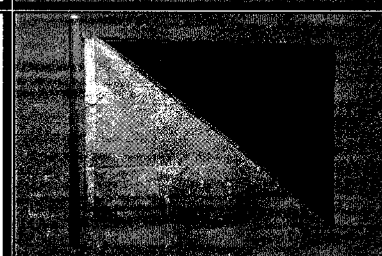
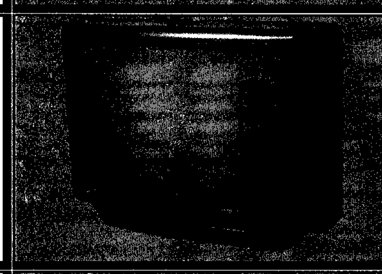
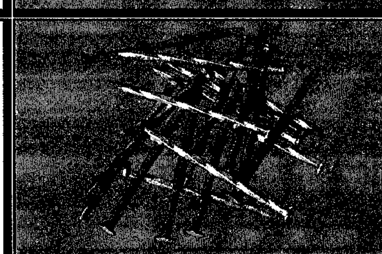
Noun	
Color	
Shape	
Semantic	
Phonological	
Unrelated	

Appendix A2

The priest *collected* the money (c)
The priest *touched* the money (p)

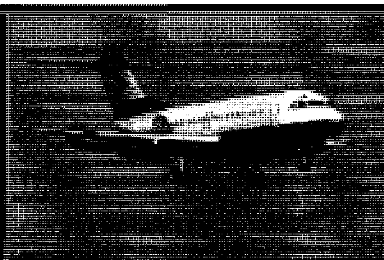
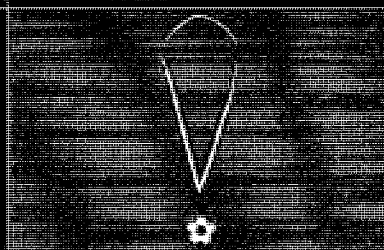
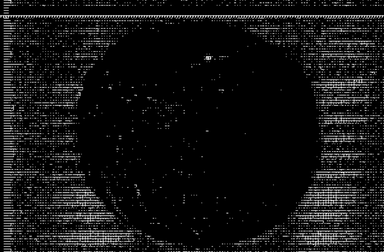
Noun	
Color	
Shape	
Semantic	
Phonological	
Unrelated	

The painter *hung* the picture (c)
The painter *eyed* the picture (p)

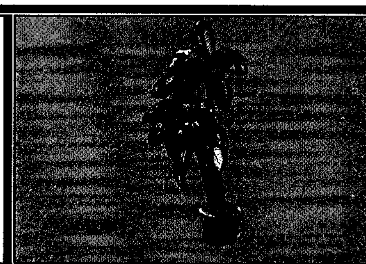
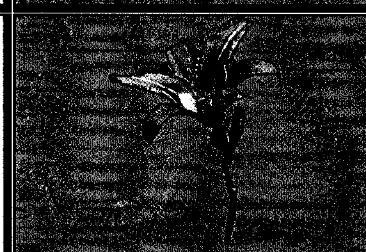
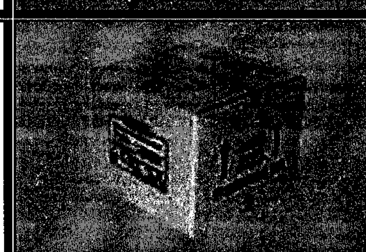
Noun	
Color	
Shape	
Semantic	
Phonological	
Unrelated	

Appendix A2

The pilot *crashed* the plane (c)
The pilot *glimpsed* the plane (p)

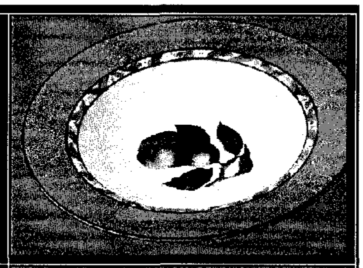
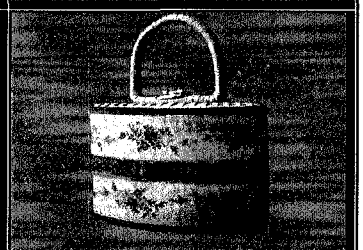
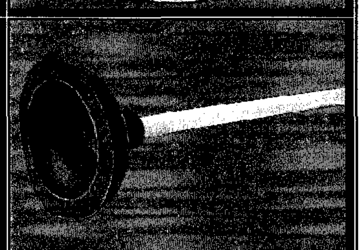
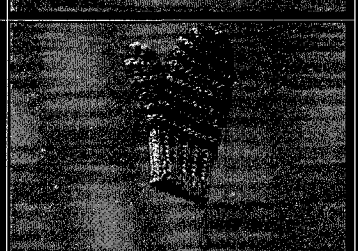
Noun	
Color	
Shape	
Semantic	
Phonological	
Unrelated	

The gardener *grew* the plant (c)
The gardener *studied* the plant (p)

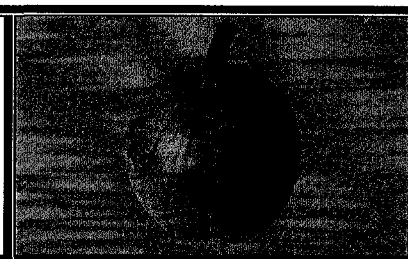
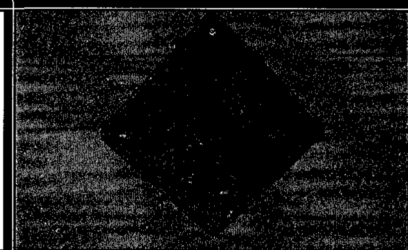
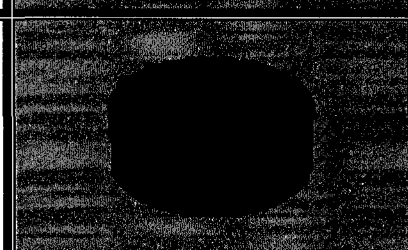
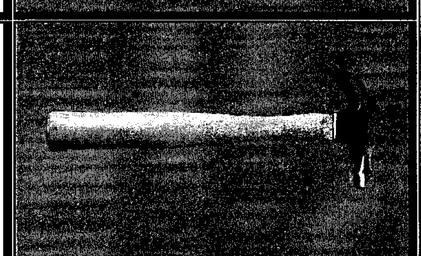
Noun	
Color	
Shape	
Semantic	
Phonological	
Unrelated	

Appendix A2

The busboy *dried* the plate (c)
The busboy *saw* the plate (p)

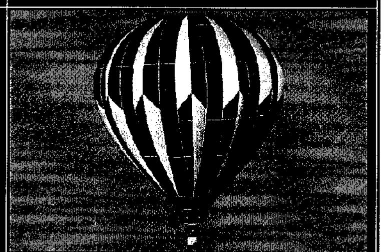
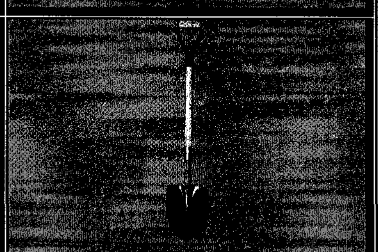
Noun	
Color	
Shape	
Semantic	
Phonological	
Unrelated	

The teacher *smashed* the pumpkin (c)
The teacher *noticed* the pumpkin (p)

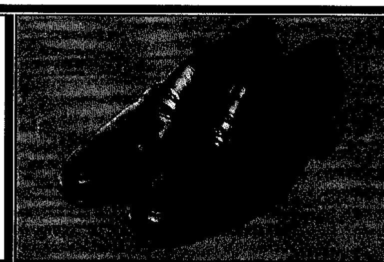
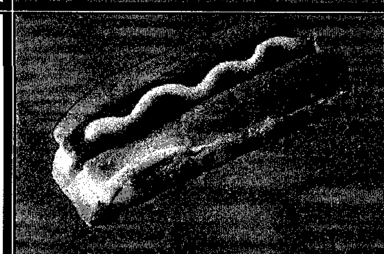
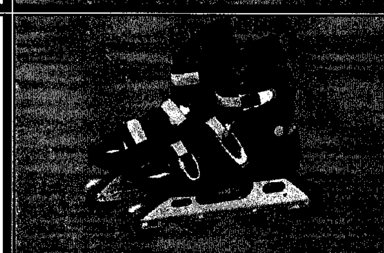
Noun	
Color	
Shape	
Semantic	
Phonological	
Unrelated	

Appendix A2

The traveler *wrinkled* the shirt (c)
The traveler *detected* the shirt (p)

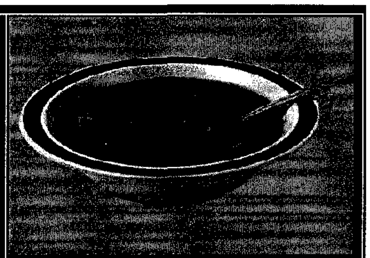
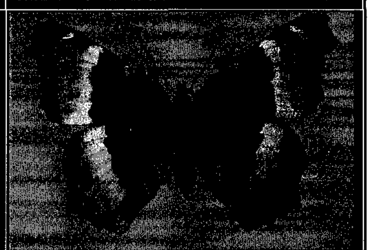
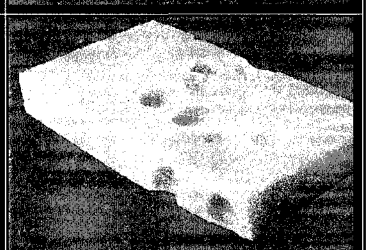
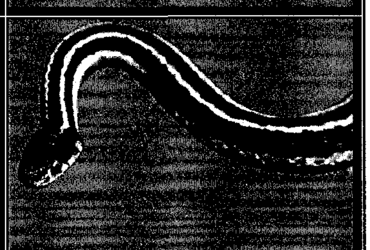
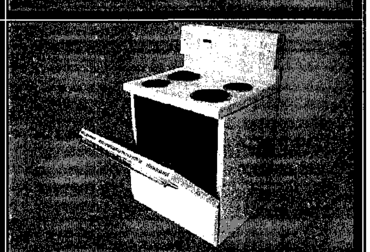
Noun	
Color	
Shape	
Semantic	
Phonological	
Unrelated	

The father *shined* the shoes (c)
The father *observed* the shoes (p)

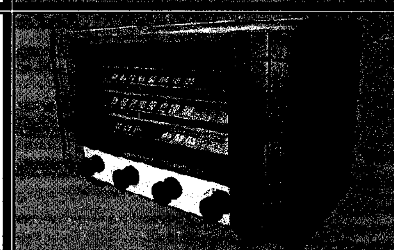
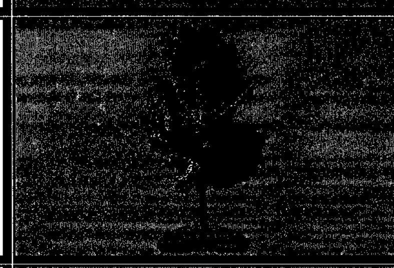
Noun	
Color	
Shape	
Semantic	
Phonological	
Unrelated	

Appendix A2

The mother *heated* the soup (c)
The mother *perceived* the soup (p)

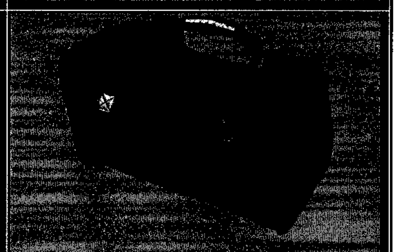
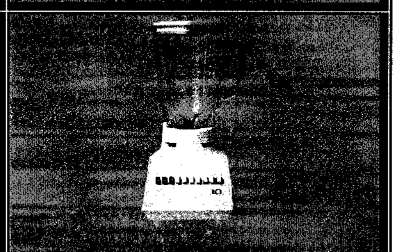
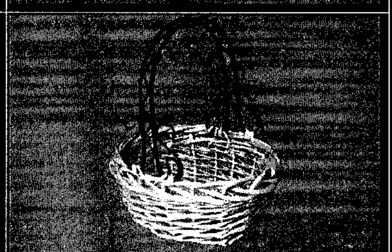
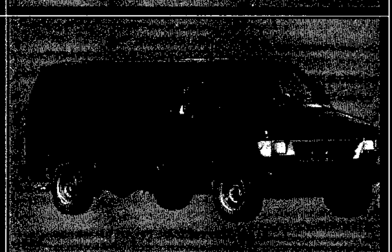
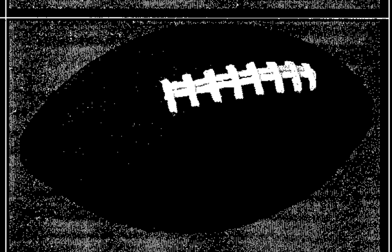
Noun	
Color	
Shape	
Semantic	
Phonological	
Unrelated	

The conductor *stopped* the train (c)
The conductor *heard* the train (p)

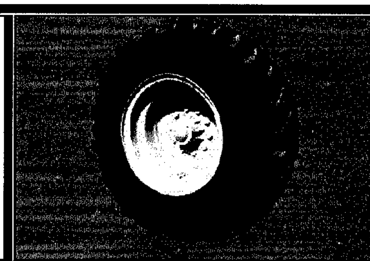
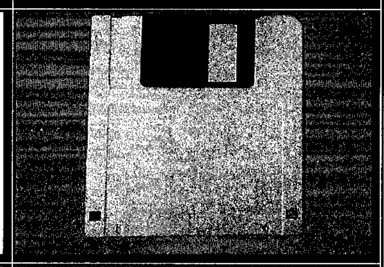
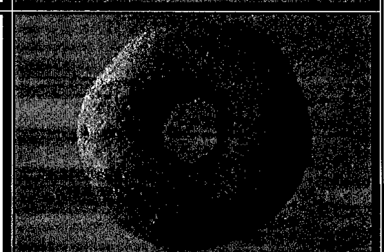
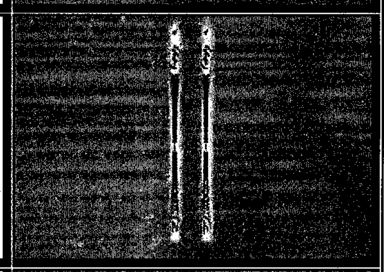
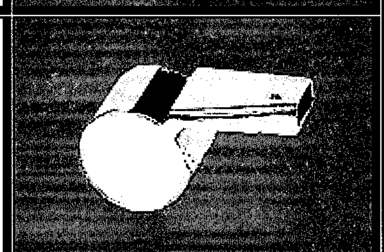
Noun	
Color	
Shape	
Semantic	
Phonological	
Unrelated	

Appendix A2

The boy *broke* the vase (c)
The boy *examined* the vase (p)

Noun	
Color	
Shape	
Semantic	
Phonological	
Unrelated	

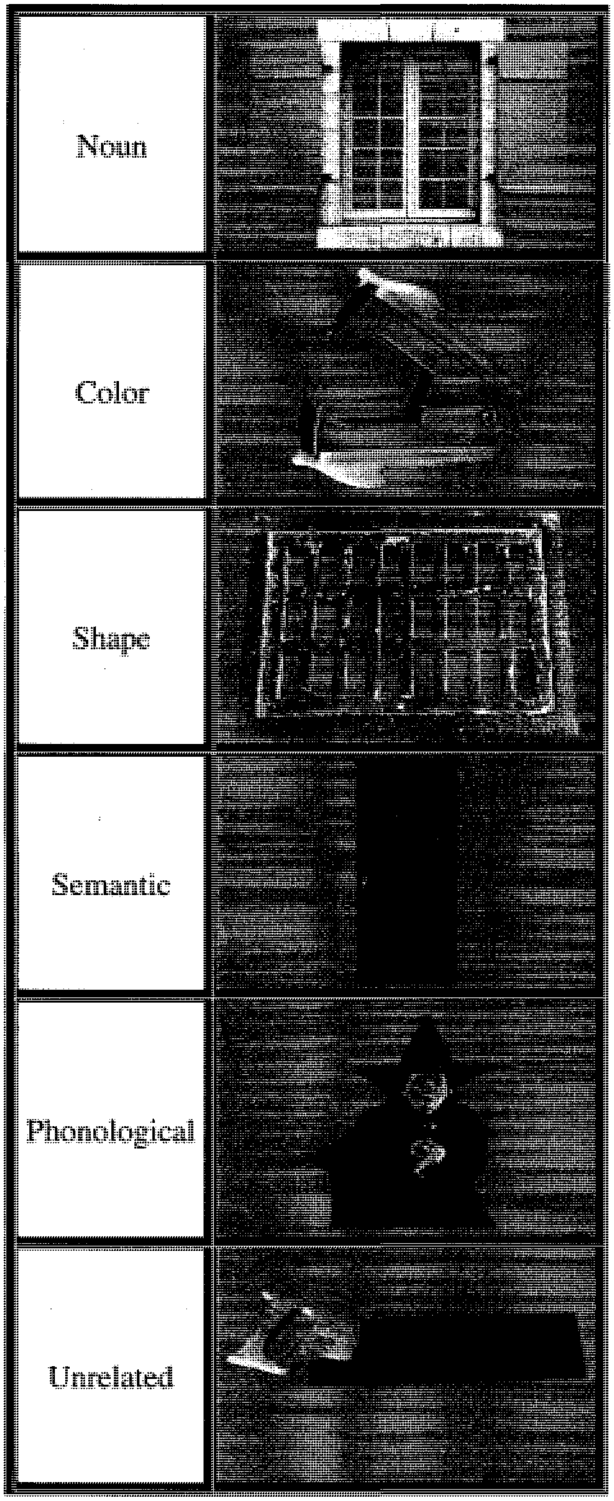
The mechanic *rotated* the wheel (c)
The mechanic *scrutinized* the wheel (p)

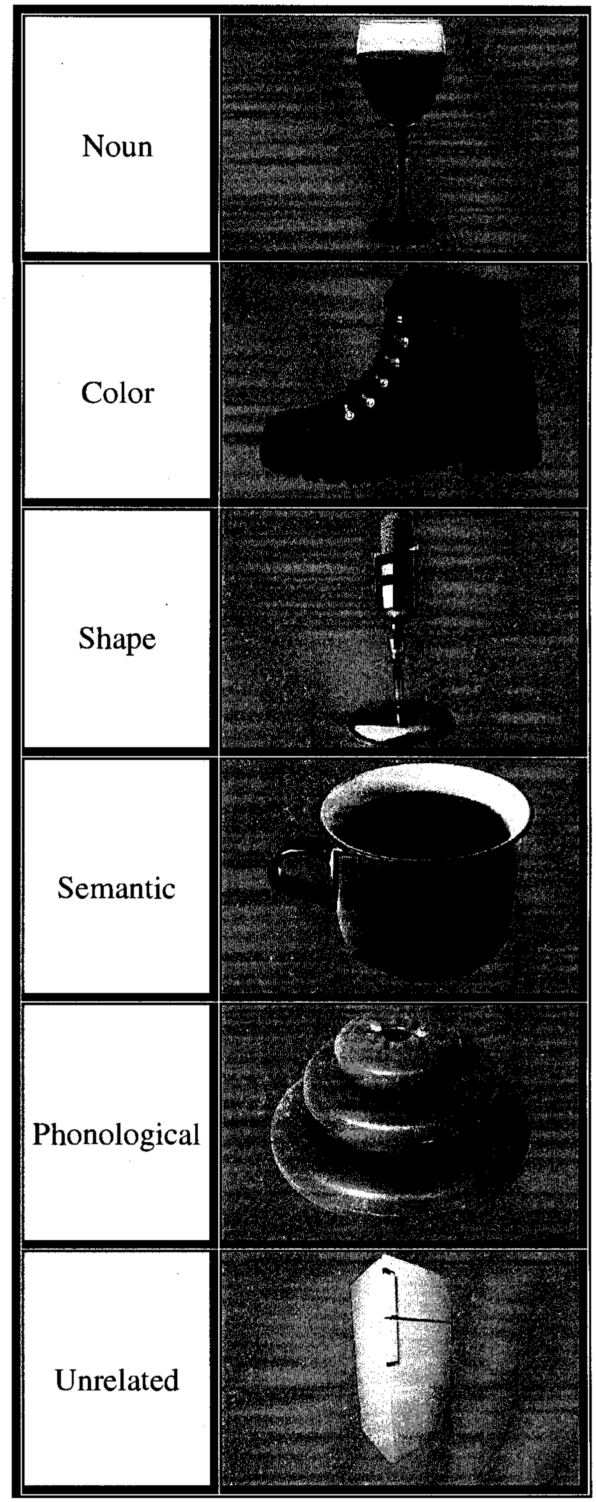
Noun	
Color	
Shape	
Semantic	
Phonological	
Unrelated	

Appendix A2

The carpenter *shattered* the window (c)
The carpenter *recognized* the window (p)

The waiter *chilled* the wine (c)
The waiter *witnessed* the wine (p)





Appendix A3

Filler Sentences

1. The child *drank* the juice
2. The farmer *ate* the sandwich
3. The boy *walked* the dog
4. The pilot *landed* the plane
5. The player *kicked* the can
6. The teenager *scratched* the CD
7. The boy *used* the skateboard
8. The student *read* the book
9. The busboy *washed* the pan
10. The mechanic *fixed* the truck
11. The baker *opened* the jar
12. The butcher *cut* the meat
13. The worker *pressed* the button
14. The teacher *corrected* the exam
15. The cook *tapped* the dough
16. The man *wrote* the poem
17. The mother *ironed* the pants
18. The mailman *closed* the package
19. The athlete ran the race
20. The father *removed* the toy
21. The secretary *typed* the letter
22. The boxer *punched* the man
23. The jeweler *rated* the diamond
24. The woman *bought* the tickets
25. The grandmother *rocked* the chair
26. The farmer *pulled* the cart
27. The mother *pushed* the carriage
28. The teenager *played* the stereo
29. The student *practiced* the guitar
30. The woman *baked* the pie
31. The waiter *moved* the table
32. The artist *made* the present
33. The designer *created* the dress
34. The boy *finished* the puzzle
35. The girl *skipped* the rope
36. The man *covered* the telephone

Appendix A4

Gender: _____ Age: _____ First Language: ______________________

Please fill in the blanks with **objects** that you believe best match the **verb** in the sentence.

1) The woman will wipe the __________ with the __________
2) The man will carve the __________ with the __________
3) The woman will mop the __________ with the __________
4) The man will shovel the __________ with the __________
5) The woman will write the __________ with the __________
6) The man will draw the __________ with the __________
7) The woman will knit the __________ with the __________
8) The woman will grate the __________ with the __________
9) The man will paint the __________ with the __________
10) The man will saw the __________ with the __________
11) The woman will snap the __________ with the __________
12) The man will hit the __________ with the __________
13) The woman will paddle the __________ with the __________
14) The woman will wash the __________ with the __________
15) The man will sweep the __________ with the __________
16) The woman will sew the __________ with the __________
17) The woman will iron the __________ with the __________
18) The man will hammer the __________ with the __________
19) The man will cut the __________ with the __________
20) The woman will chop the __________ with the __________
21) The man will drill the __________ with the __________
22) The woman will chip the __________ with the __________
23) The woman will scrape __________ with the __________
24) The man will chisel the __________ with the __________
25) The man will snip the __________ with the __________
26) The woman will bore the __________ with the __________
27) The man will crush the __________ with the __________
28) The woman will dice the __________ with the __________
29) The man will file the __________ with the __________
30) The man will mash the __________ with the __________
31) The woman will mince the __________ with the __________
32) The man will clip the __________ with the __________
33) The man will slice the __________ with the __________
34) The woman will squash the __________ with the __________
35) The man will record the __________ with the __________
36) The woman will rake the __________ with the __________
37) The man will type the __________ with the __________
38) The man will scramble the __________ with the __________
39) The woman will toss the __________ with the __________
40) The man will brush the __________ with the __________

Appendix A4

41) The man will fax the __________ with the __________
42) The man will sponge the __________ with the __________
43) The woman will stir the __________ with the __________
44) The woman will beat the __________ with the __________
45) The man will dust the __________ with the __________
46) The woman will comb the __________ with the __________
47) The woman will polish the __________ with the __________
48) The man will hose the __________ with the __________
49) The woman will cover the __________ with the __________
50) The man will dry the __________ with the __________
51) The woman will erase the __________ with the __________
52) The man will filter the __________ with the __________
53) The woman will trim the __________ with the __________
54) The man will whisk the __________ with the __________
55) The man will staple the __________ with the __________
56) The woman will rub the __________ with the __________
57) The man will wax the __________ with the __________
58) The woman will scrub the __________ with the __________
59) The woman will vacuum the __________ with the __________
60) The man will shave the __________ with the __________
61) The woman will perforate the __________ with the __________
62) The man will light the __________ with the __________

Appendix A5

Pilot Movie Scenarios

1) Immediately after reading the recipe the woman will chop the vegetables with the knife.
2) Right after getting the soda the woman will chip the ice with the pick.
3) Later, after pouring the flour the man will whisk the eggs in the dining room.
4) Immediately after drying his hands the man will carve the turkey with the knife.
5) Right after storing the leftovers the woman will wash the dishes with the cloth.
6) Later, after draining the pasta the woman will grate the cheese in the kitchen.
7) Immediately after melting the butter the woman will dice the onions with the knife.
8) Right after preparing the coffee the man will scramble the eggs with the fork.
9) Later, after turning on the oven the man will mash the potatoes in the kitchen.
10) Immediately after cooking the bacon the man will slice the bread with the knife.
11) Right after heating the milk the woman will beat the eggs in the kitchen.
12) Later, after folding the cloth the woman will iron the shirt in the kitchen.
13) Immediately after fluffing the pillows the man will dust the table in the living room.
14) Right after arranging the books the woman will wipe the table with the cloth.
15) Later, after setting the table the woman will toss the salad with the tongs.
16) Immediately after sipping the wine the woman will stir the soup with the spoon.
17) Right after checking the wine the man will light the candle with the match.
18) Later, after doing the laundry the woman will vacuum the carpet in the basement.
19) Immediately after reading the book the woman will knit the sweater with the needles.
20) Right after checking the e-mails the man will cut the paper with the scissors.
21) Later, after talking on the phone the woman will write the letter with the pen.
22) Immediately after getting the file the man will fax the letter in the office.
23) Right after working on the computer the woman will erase the board in the office.
24) Later, after reading the textbook the man will staple the papers in the office.
25) Immediately after wiping the home plate the man will hit the ball with the bat.
26) Right after studying the plans the man will saw the wood in the backyard.
27) Later, after inspecting the life jacket the woman will paddle the boat in the lake.
28) Immediately after searching the tool box the man will hammer the nail in the backyard.
29) Right after drawing the sketch the man will chisel the wood in the backyard.
30) Later, after watering the plant the woman will rake the leaves in the garden.
31) Immediately after adjusting the camera the man will brush the dog in the garden.
32) Right after leafing the book the man will draw the picture with the pencil.

Appendix A6

You will watch a series of movies. In some of them there will be an object that will disappear. Please rate from a scale of 1 to 5 how noticeable this disappearance is. With 1 indicating that the disappearance was very noticeable and 5 indicating that the disappearance was not noticeable.

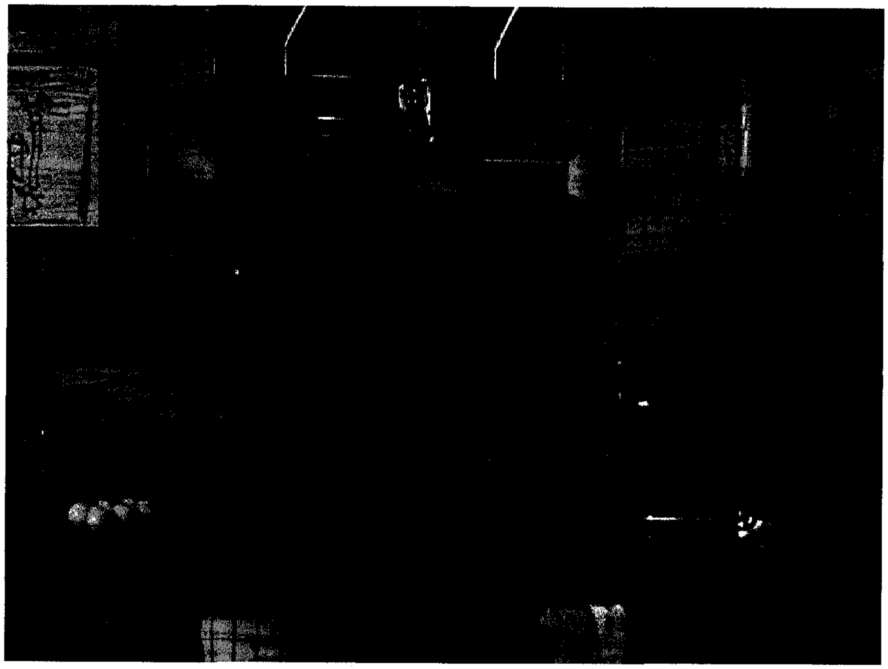
	Very Noticeable		Somewhat Noticeable		Not Noticeable
1)	1	2	3	4	5
2)	1	2	3	4	5
3)	1	2	3	4	5
4)	1	2	3	4	5
5)	1	2	3	4	5
6)	1	2	3	4	5
7)	1	2	3	4	5
8)	1	2	3	4	5
9)	1	2	3	4	5
10)	1	2	3	4	5
11)	1	2	3	4	5

-- same format up to 96

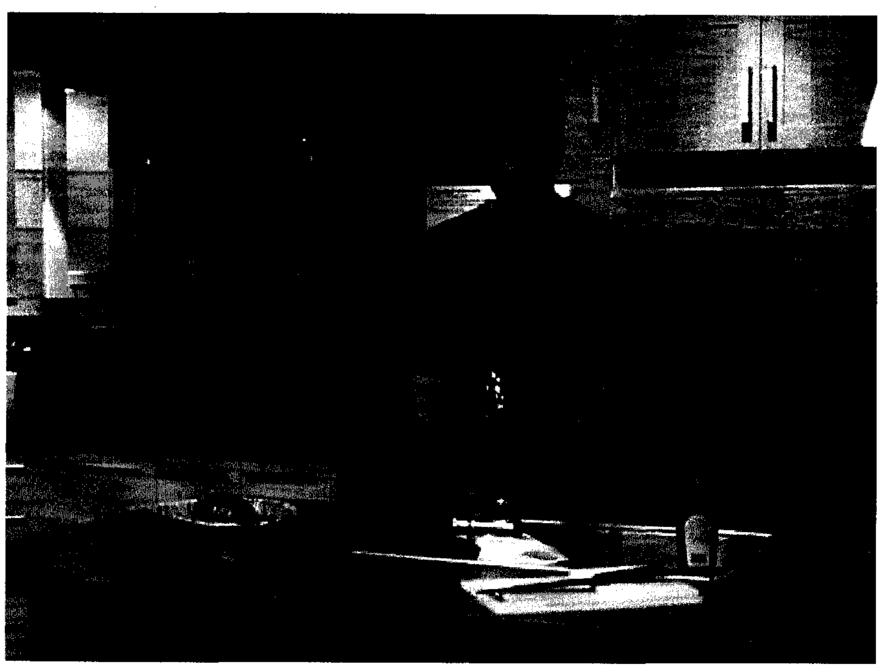
96)	1	2	3	4	5

Appendix A7

Right after heating the milk the woman will *beat* the eggs in the kitchen (d)

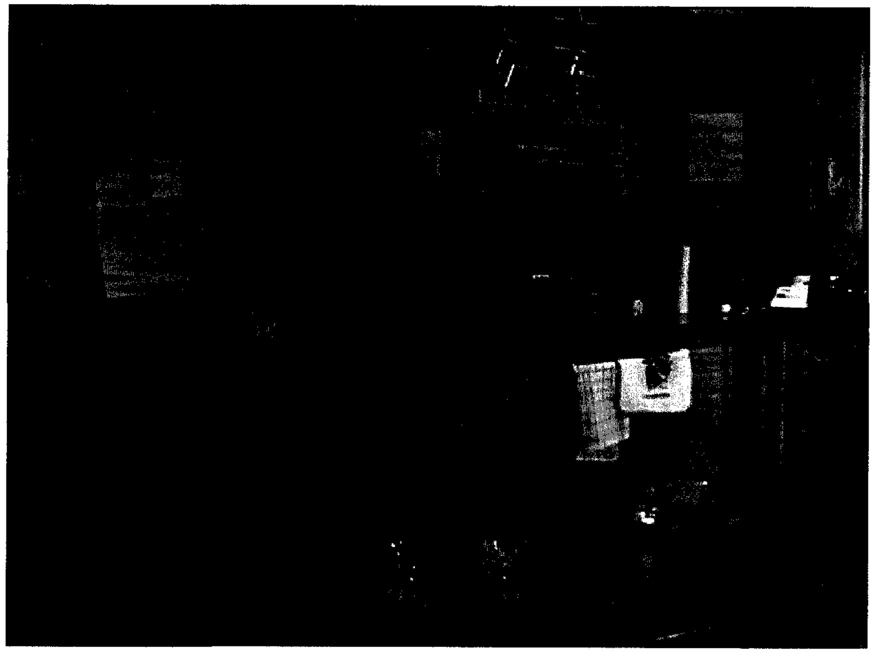


Immediately after washing his hands the man will *carve* the turkey with the knife (nd)

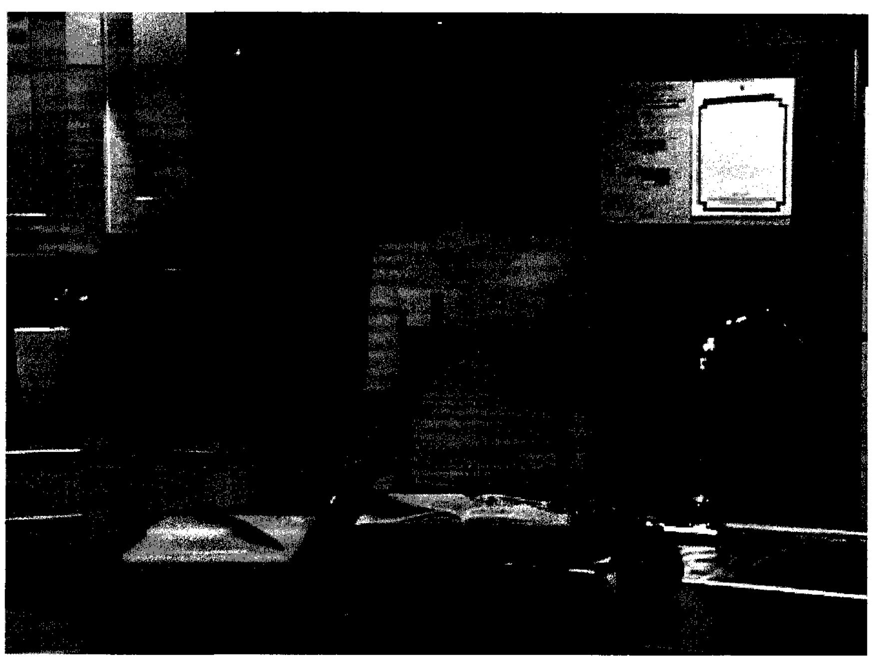


Appendix A7

Right after getting the soda bottle the woman will *chip* the ice with the pick (nd)



Immediately after reading the recipe the woman will *chop* the vegetables with the knife (nd)



Appendix A7

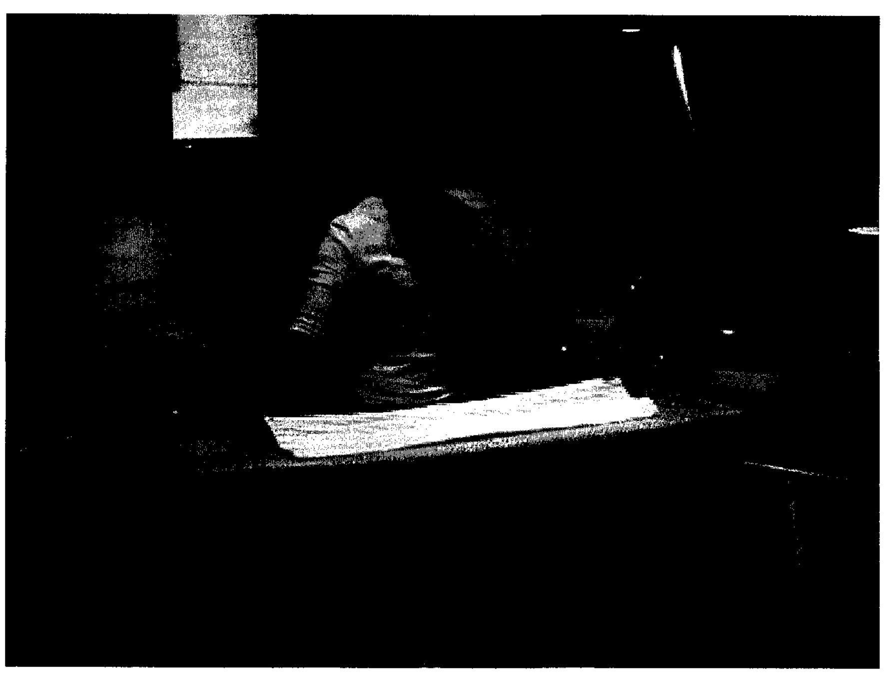
Right after checking his emails the man will *cut* the paper with the scissors (nd)

Immediately after fluffing the pillows the man will *dust* the table in the living room (d)

Appendix A7

Immediately after searching the tool box the man will *hammer* the nail in the backyard(d)

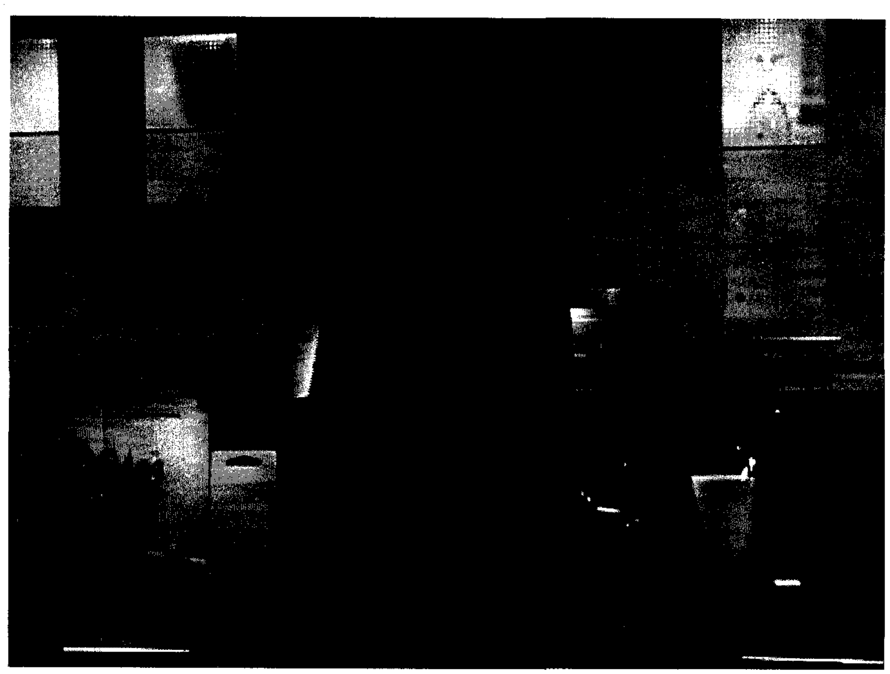
Later, after folding the cloth the woman will *iron* the shirt in the kitchen (d)



Appendix A7

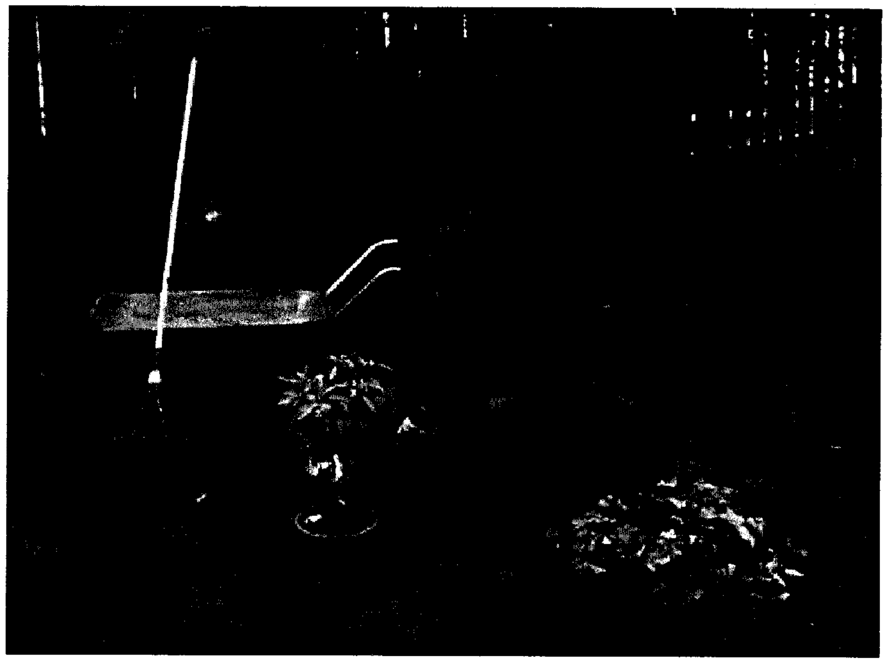
Right after checking the wine the man will *light* the candle with the match (nd)

Later, after turning on the oven the man will *mash* the potatoes in the kitchen (d)

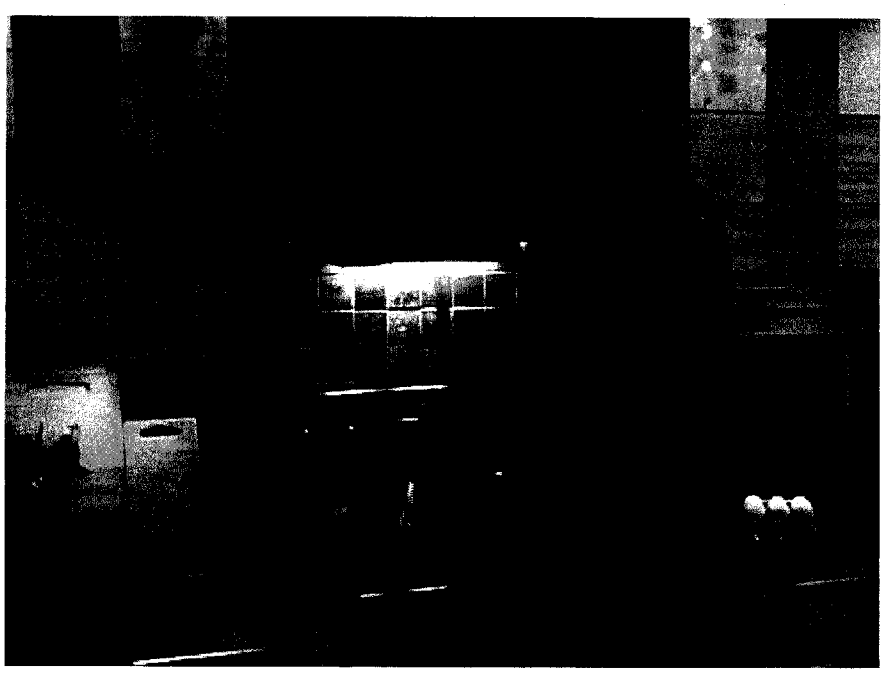


Appendix A7

Later, after watering the plant the woman will *rake* the leaves in the garden (d)

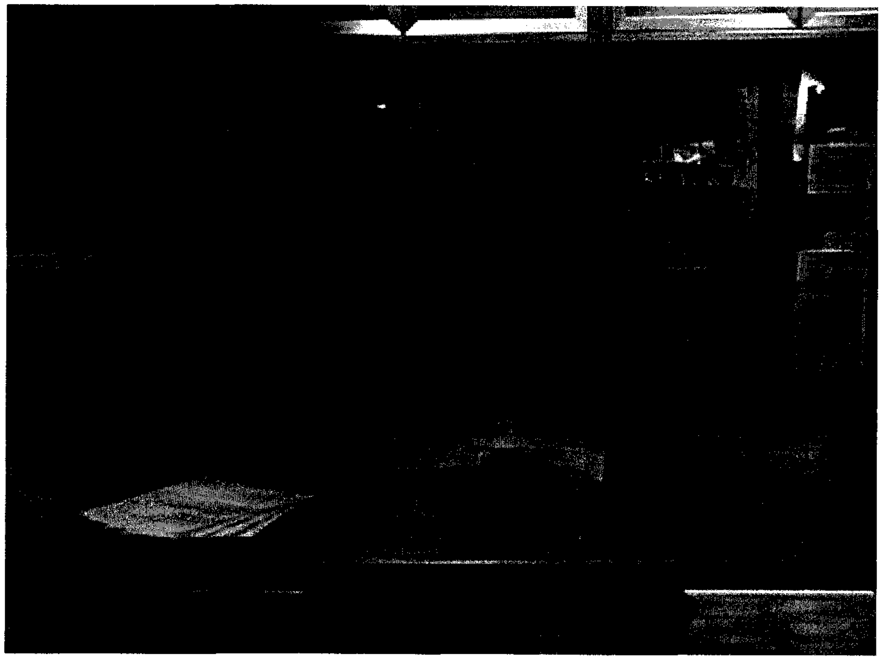


Right after preparing the coffee the man will *scramble* the eggs with the fork (nd)



Appendix A7

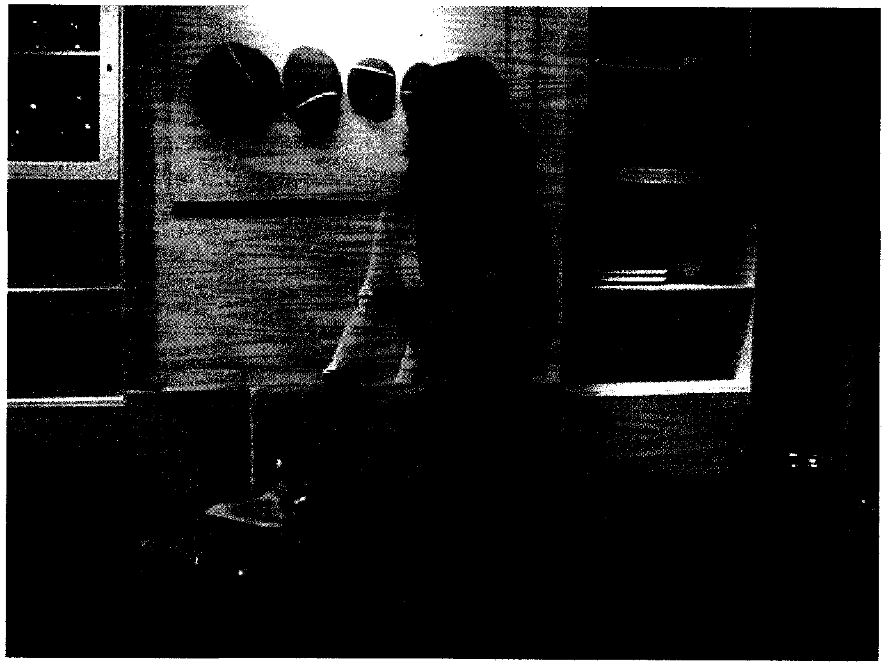
Later, after reading the textbook the man will *staple* the papers in the office (d)



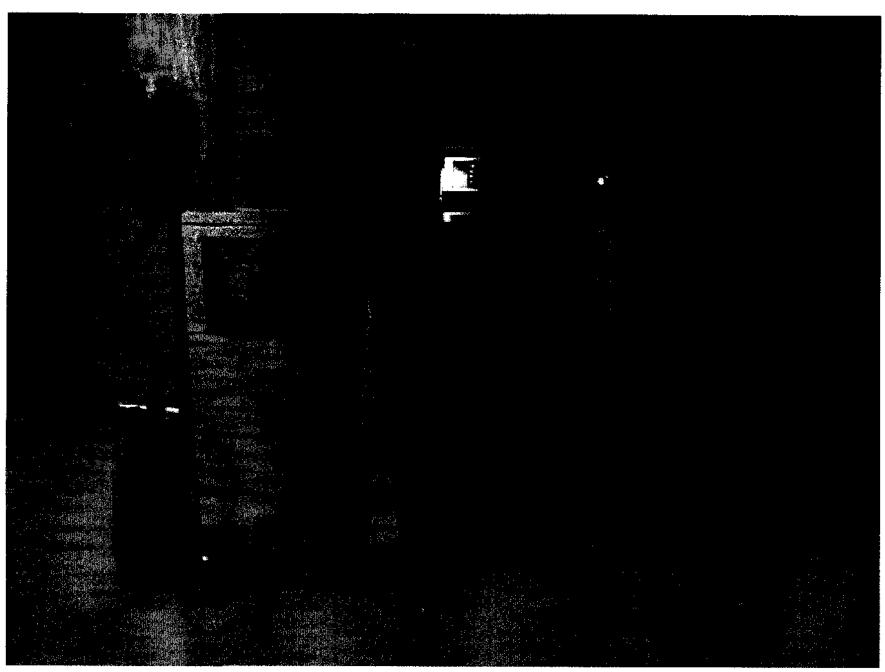
Immediately after sipping the wine the woman will *stir* the soup with the spoon (nd)

Appendix A7

Later, after setting the table the woman will *toss* the salad with the thongs (nd)



Later, after doing the laundry the woman will *vacuum* the carpet in the basement (d)



Appendix A7

Right after storing the leftovers the woman will *wash* the dishes with the cloth (nd)

Later, after pouring the flour the man will *whisk* the eggs in the dining room (d)

Appendix A8

You will watch a series of pictures. Please indicate by "yes" or "no" whether the object displayed was present in any of the movies you just saw, and identify the object.

1)	Yes	No	____________________
2)	Yes	No	____________________
3)	Yes	No	____________________
4)	Yes	No	____________________
5)	Yes	No	____________________
6)	Yes	No	____________________
7)	Yes	No	____________________
8)	Yes	No	____________________
9)	Yes	No	____________________
10)	Yes	No	____________________
11)	Yes	No	____________________
12)	Yes	No	____________________

-- same format up to 72

72)	Yes	No	____________________

Appendix A9

Filler Sentences

1. Later, after draining the pasta the woman will *grate* the cheese in the kitchen (d)
2. Immediately after melting the butter the woman will *dice* the onions with the knife (nd)
3. Right after arranging the books the woman will *wipe* the table with the cloth (nd)
4. Immediately after reading the book the woman will *knit* the sweater with the needles (nd)
5. Later, after talking on the phone the woman will *write* the letter with the pen (nd)
6. Immediately after getting the file the man will *fax* the letter in the office (d)
7. Right after working on the computer the woman will *erase* the board in the office (d)
8. Immediately after wiping the home plate the man will *hit* the ball with the bat (nd)
9. Right after studying the plans the man will *saw* the wood in the backyard (d)
10. Later, after inspecting the life jacket the woman will *paddle* the boat in the lake (d)
11. Right after drawing the sketch the man will *chisel* the wood in the backyard (d)
12. Right after leafing the book the man will *draw* the picture with the pencil (nd)

Appendix B

Consent Forms Used for Normative and Pilot Studies

and Experiments 1a, 1b, 1c, 2a and 2b

Appendix B1

CONSENT FORM TO PARTICIPATE IN RESEARCH

This is to state that I agree to participate in a program of research being conducted by Caroline van de Velde of the Department of Psychology at Concordia University, in conjunction with her Doctorate Thesis project, under the supervision of Dr. Michael von Grünau.

A. PURPOSE

I have been informed that the purpose of the research is to investigate how the brain processes visual information and it's interaction with language. This is being conducted to partially fulfill the requirements for a Doctoral Thesis.

B. PROCEDURES

I have been informed that the experiment involves the following procedures: I will be required to look at pairs of pictures of objects and rate them from I to 5 - whereby "1" designated "not at all similar" and "5" designated "very similar" – on how similar they are in one of 4 features (name, function, color and shape). The completion of the experiment will take about 20 minutes, and in return for my participation I will receive 1 course credit. I have been informed that my name will not be associated with my data in the experiment. I understand that my participation in the experiment, and the information and data I provide, will be kept strictly confidential. I understand that if the results are published, individual data will be reported but I will be only indicated as "Participant" and then a number.

C. CONDITIONS OF PARTICIPATION

- I understand that I am free to decline to participate in the experiment without negative consequences.
- I understand that I am free to withdraw my consent and discontinue my participation at anytime without negative consequences.
- I understand that my participation in this study is confidential (i.e., the researcher will know, but my identity will not be disclosed).
- I understand that my data from this research may be published.

I HAVE CAREFULLY STUDIED THE ABOVE AND UNDERSTAND THIS AGREEMENT. I FREELY CONSENT AND AGREE TO PARTICIPATE IN THIS STUDY.

NAME (please print) ____

SIGNATURE ____

WITNESS SIGNATURE ____

DATE _____________________

If at any time you have questions about your rights as a research participant, please contact Adela Ried, Compliance Officer, Concordia University, at 514.290.0523 or by email at areid@alcor.concordia.ca.

Appendix B2

CONSENT FORM TO PARTICIPATE IN RESEARCH

This is to state that I agree to participate in a program of research being conducted by Caroline van de Velde of the Department of Psychology at Concordia University, in conjunction with her Doctorate Thesis project, under the supervision of Dr. Michael von Grünau.

A. PURPOSE

I have been informed that the purpose of the research is to investigate how the brain processes visual information and it's interaction with language. This is being conducted to partially fulfill the requirements for a Doctorate Thesis.

B. PROCEDURES

I have been informed that the experiment involves the following procedures: An eye-tracker machine will be placed over my head that will record my eye movements as I look at pictures of objects on a computer screen. I will also hear sentences through headphones. A central dot will be present, when it is red I am not allowed to move my eyes but must keep fixated on the red dot, when it is green it means I am free to move my eyes. A set of six pictures will appear briefly on the screen and my task will be to scan each one until they disappear. When the objects reappear, my task will be to indicate whether the picture of the object uttered in the sentence is present (green key if present, red key if not present) and to look at the pictures of the object if it is present before responding. In order to initiate each trial I will be required to press the yellow key on the button box.

The completion of the experiment will take about 30 minutes. I have been informed that my name will not be associated with my data in the experiment. I understand that my participation in the experiment, and the information and data I provide, will be kept strictly confidential. I understand that if the results are published, individual data will be reported but I will be only indicated as "Participant" and then a number.

C. CONDITIONS OF PARTICIPATION

- I understand that I am free to decline to participate in the experiment without negative consequences.
- I understand that I am free to withdraw my consent and discontinue my participation at anytime without negative consequences.
- I understand that my participation in this study is confidential (i.e., the researcher will know, but my identity will not be disclosed).
- I understand that my data from this research may be published.
- I understand the purpose of this study and know there is no hidden motive of which I have not been informed.

I HAVE CAREFULLY STUDIED THE ABOVE AND UNDERSTAND THIS AGREEMENT. I FREELY CONSENT AND AGREE TO PARTICIPATE IN THIS STUDY.

NAME (please print) ______________________________

SIGNATURE ______________________________

WITNESS SIGNATURE ______________________________

DATE ____________________

If at any time you have questions about your rights as a research participant, please contact Adela Ried, Compliance Officer, Concordia University, at 514.290.0523 or by email at areid@alcor.concordia.ca.

Appendix B3

CONSENT FORM TO PARTICIPATE IN RESEARCH

This is to state that I agree to participate in a program of research being conducted by Caroline van de Velde of the Department of Psychology at Concordia University, in conjunction with her Doctorate Thesis project, under the supervision of Dr. Michael von Grünau.

A. PURPOSE

I have been informed that the purpose of the research is to investigate how the brain processes visual information and it's interaction with language. This is being conducted to partially fulfill the requirements for a Doctoral Thesis.

B. PROCEDURES

I have been informed that the experiment involves the following procedures: I will be required to fill in the blanks with objects that best fit the context of the sentence. The completion of the experiment will take about 20 minutes, and in return for my participation I will receive 1 course credit. I have been informed that my name will not be associated with my data in the experiment. I understand that my participation in the experiment, and the information and data I provide, will be kept strictly confidential. I understand that if the results are published, individual data will be reported but I will be only indicated as "Participant" and then a number.

C. CONDITIONS OF PARTICIPATION

- I understand that I am free to decline to participate in the experiment without negative consequences.
- I understand that I am free to withdraw my consent and discontinue my participation at anytime without negative consequences.
- I understand that my participation in this study is confidential (i.e., the researcher will know, but my identity will not be disclosed).
- I understand that my data from this research may be published.

I HAVE CAREFULLY STUDIED THE ABOVE AND UNDERSTAND THIS AGREEMENT. I FREELY CONSENT AND AGREE TO PARTICIPATE IN THIS STUDY.

NAME (please print) ____

SIGNATURE ____

WITNESS SIGNATURE ____

DATE ______________________

If at any time you have questions about your rights as a research participant, please contact Adela Ried, Compliance Officer, Concordia University, at 514.290.0523 or by email at areid@alcor.concordia.ca.

Appendix B4

CONSENT FORM TO PARTICIPATE IN RESEARCH

This is to state that I agree to participate in a program of research being conducted by Caroline van de Velde of the Department of Psychology at Concordia University, in conjunction with her Doctorate Thesis project, under the supervision of Dr. Michael von Grünau.

A. PURPOSE

I have been informed that the purpose of the research is to investigate how the brain processes visual information and it's interaction with language. This is being conducted to partially fulfill the requirements for a Doctoral Thesis.

B. PROCEDURES

I have been informed that the experiment involves the following procedures: I will be required to look at short movies and after each trial describe in detail what I just saw in the movie. The completion of the experiment will take about 20 minutes, and in return for my participation I will receive $5.00. I have been informed that my name will not be associated with my data in the experiment. I understand that my participation in the experiment, and the information and data I provide, will be kept strictly confidential. I understand that if the results are published, individual data will be reported but I will be only indicated as "Participant" and then a number.

C. CONDITIONS OF PARTICIPATION

- I understand that I am free to decline to participate in the experiment without negative consequences.
- I understand that I am free to withdraw my consent and discontinue my participation at anytime without negative consequences.
- I understand that my participation in this study is confidential (i.e., the researcher will know, but my identity will not be disclosed).
- I understand that my data from this research may be published.

I HAVE CAREFULLY STUDIED THE ABOVE AND UNDERSTAND THIS AGREEMENT. I FREELY CONSENT AND AGREE TO PARTICIPATE IN THIS STUDY.

NAME (please print) ____

SIGNATURE ____

WITNESS SIGNATURE ____

DATE ______________________

If at any time you have questions about your rights as a research participant, please contact Adela Ried, Compliance Officer, Concordia University, at 514.290.0523 or by email at areid@alcor.concordia.ca.

Appendix B5

CONSENT FORM TO PARTICIPATE IN RESEARCH

This is to state that I agree to participate in a program of research being conducted by Caroline van de Velde of the Department of Psychology at Concordia University, in conjunction with her Doctorate Thesis project, under the supervision of Dr. Michael von Grünau.

A. PURPOSE

I have been informed that the purpose of the research is to investigate how the brain processes visual information and it's interaction with language. This is being conducted to partially fulfill the requirements for a Doctoral Thesis.

B. PROCEDURES

I have been informed that the experiment involves the following procedures: I will be required to look at a set of short movies where in some an item disappears. After each trial I will be asked to rate from a scale of 1 to 5 how noticeable the disappearance was – with "1" indicating that the disappearance was very noticeable and "5" indicating that the disappearance was not noticeable. The completion of the experiment will take about 30 minutes, and in return for my participation I will receive 1 course credit. I have been informed that my name will not be associated with my data in the experiment. I understand that my participation in the experiment, and the information and data I provide, will be kept strictly confidential. I understand that if the results are published, individual data will be reported but I will be only indicated as "Participant" and then a number.

C. CONDITIONS OF PARTICIPATION

- I understand that I am free to decline to participate in the experiment without negative consequences.
- I understand that I am free to withdraw my consent and discontinue my participation at anytime without negative consequences.
- I understand that my participation in this study is confidential (i.e., the researcher will know, but my identity will not be disclosed).
- I understand that my data from this research may be published.

I HAVE CAREFULLY STUDIED THE ABOVE AND UNDERSTAND THIS AGREEMENT. I FREELY CONSENT AND AGREE TO PARTICIPATE IN THIS STUDY.

NAME (please print) ____

SIGNATURE ____

WITNESS SIGNATURE ____

DATE ______________________

If at any time you have questions about your rights as a research participant, please contact Adela Ried, Compliance Officer, Concordia University, at 514.290.0523 or by email at areid@alcor.concordia.ca.

Appendix B6

CONSENT FORM TO PARTICIPATE IN RESEARCH

This is to state that I agree to participate in a program of research being conducted by Caroline van de Velde of the Department of Psychology at Concordia University, in conjunction with her Doctorate Thesis project, under the supervision of Dr. Michael von Grünau.

A. PURPOSE

I have been informed that the purpose of the research is to investigate how the brain processes visual information and it's interaction with language. This is being conducted to partially fulfill the requirements for a Doctoral Thesis.

B. PROCEDURES

I have been informed that the experiment involves the following procedures: I will be required to look at short movies while my eye movements are being recorded with eye tracker equipment. After each trial I will be asked to describe in detail, to the experimenter, what I just saw in the short movies. The completion of the experiment will take about 1 hour, and in return for my participation I will receive 1 course credit or $10.00. I have been informed that my name will not be associated with my data in the experiment. I understand that my participation in the experiment, and the information and data I provide, will be kept strictly confidential. I understand that if the results are published, individual data will be reported but I will be only indicated as "Participant" and then a number.

C. CONDITIONS OF PARTICIPATION

- I understand that I am free to decline to participate in the experiment without negative consequences.
- I understand that I am free to withdraw my consent and discontinue my participation at anytime without negative consequences.
- I understand that my participation in this study is confidential (i.e., the researcher will know, but my identity will not be disclosed).
- I understand that my data from this research may be published.

I HAVE CAREFULLY STUDIED THE ABOVE AND UNDERSTAND THIS AGREEMENT. I FREELY CONSENT AND AGREE TO PARTICIPATE IN THIS STUDY.

NAME (please print) ____

SIGNATURE ____

WITNESS SIGNATURE ____

DATE ________________________

If at any time you have questions about your rights as a research participant, please contact Adela Ried, Compliance Officer, Concordia University, at 514.290.0523 or by email at areid@alcor.concordia.ca.

Appendix B7

CONSENT FORM TO PARTICIPATE IN RESEARCH

This is to state that I agree to participate in a program of research being conducted by Caroline van de Velde of the Department of Psychology at Concordia University, in conjunction with her Doctorate Thesis project, under the supervision of Dr. Michael von Grünau.

A. PURPOSE

I have been informed that the purpose of the research is to investigate how the brain processes visual information and it's interaction with language. This is being conducted to partially fulfill the requirements for a Doctoral Thesis.

B. PROCEDURES

I have been informed that the experiment involves the following procedures: I will be required to look at short movies while my eye movements are being recorded with eye tracker equipment. During the presentation of each movie I will hear a sentence through headphones describing the events in the movie. After each trial I will be asked to describe in detail, to the experimenter, what I just saw in the short movies. The completion of the experiment will take about 1 hour, and in return for my participation I will receive 1 course credit or $10.00. I have been informed that my name will not be associated with my data in the experiment. I understand that my participation in the experiment, and the information and data I provide, will be kept strictly confidential. I understand that if the results are published, individual data will be reported but I will be only indicated as "Participant" and then a number.

C. CONDITIONS OF PARTICIPATION

- I understand that I am free to decline to participate in the experiment without negative consequences.
- I understand that I am free to withdraw my consent and discontinue my participation at anytime without negative consequences.
- I understand that my participation in this study is confidential (i.e., the researcher will know, but my identity will not be disclosed).
- I understand that my data from this research may be published.

I HAVE CAREFULLY STUDIED THE ABOVE AND UNDERSTAND THIS AGREEMENT. I FREELY CONSENT AND AGREE TO PARTICIPATE IN THIS STUDY.

NAME (please print) ____

SIGNATURE ____

WITNESS SIGNATURE ____

DATE ________________________

If at any time you have questions about your rights as a research participant, please contact Adela Ried, Compliance Officer, Concordia University, at 514.290.0523 or by email at areid@alcor.concordia.ca.

Appendix C

Instructions Used for Normative and Pilot Studies and Experiments 1a, 1b, 1c, 2a and 2b

Appendix C1

Instructions for Experiment 1a, 1b, and 1c

In this experiment you will see several pictures of objects displayed on the screen. At a certain point you will hear a sentence that refers to one of the objects. Your task will be to look at the object that the sentence refers to as quickly as possible.

There are several events before you get the sentence that matches the object. Please read carefully the sequence of events and what you should do in each part of the experimental trials.

1- First, each trial will start with an asterisk (*) displayed in the middle of the screen. You should keep your eyes on this asterisk and press the YELLOW key when you are ready to begin the new trial.

2- Second, you will see a set of 6 pictures of objects presented on the screen. The first time these pictures appear you will also see a GREEN dot where the asterisk was displayed. The GREEN dot means that you are FREE to move your eyes and look at all the pictures displayed on the screen. At this point it is important that you look at all of the pictures because you will be required to remember one of the pictures of the display.

3- A couple of seconds after the GREEN dot will turn RED and the pictures will disappear. This means that you should immediately look at the center for the screen again, that is, that you should look at the RED dot. In order to help you notice that the dot went from green to red, you will hear a beep. Remember, at the beep the dot turns RED and you should look at it until it turns green again.

4- A couple of seconds after the dot turns RED, you will hear a sentence. Shortly after, the RED dot will turn GREEN, and the pictures will reappear. Your task is to move your eyes to the object that the sentence refers to.

5- Once you have looked at the object, you press the RED key if the object mentioned in the sentence is NOT PRESENT and press the GREEN key if the object is PRESENT. This will also start a new trial.

Appendix C2

Reinforce Instructions 1a, 1b, and 1c

So, to sum up, here's the sequence of events:

1-you see the asterisk, you press the YELLOW key to start a trial

2-you keep your eyes on the RED dot in the center of the screen

3-shortly after you will hear a sentence and then the RED dot will turn GREEN and 6 pictures of objects will appear on the screen. Your task is to look at the picture of the object mentioned in the sentence.

4-once you have looked directly at the object, press the RED key if the object mentioned in the sentence is NOT PRESENT and press the GREEN key if the object is PRESENT. This will also move you to the next trial.

Here are a couple of important things to remember:

-GREEN means MOVE YOUR EYES (to the object that the sentence names)

-RED means DON'T MOVE YOUR EYES and look at the dot

-When you are free to move, after you heard the sentence, look at the object immediately and only press the RED or GREEN key to move to the next trial once you have looked at the object

Appendix C3

Instructions for Experiment 2a

Thank you for choosing to participate in this experiment.

You will be presented with a series of movies. You are asked to simply look at the movies. Prior to each trial, there will be a fixation cross (+) in the middle of the screen that you must fixate on. This cross will be red on a black background. The movie will then begin, and the + will disappear, and you may look wherever you like on the screen. After each movie you will then be asked to describe to the experimenter in 15 seconds what you just saw in the movie.

If at any time you experience discomfort, you may choose to discontinue the experiment.

Now sit back, relax, and enjoy!

Press the spacebar to continue.

Appendix C4

Reinforce Instructions for Experiment 2a and 2b

Please describe in detail, to the experimenter, the short movie you just saw.

Appendix C5

Instructions for Experiment 2b

Thank you for choosing to participate in this experiment.

You will be presented with a series of movies. You are asked to simply look at the movies.
Prior to each trial, there will be a fixation cross (+) in the middle of the screen that you must fixate on. This cross will be red on a black background. The movie will then begin, and the + will disappear, and you may look wherever you like on the screen.
While you are watching the movie, you will hear a sentence through the headphones.
After each movie you will then be asked to describe to the experimenter in 15 seconds what you just saw in the movie.

If at any time you experience discomfort, you may choose to discontinue the experiment.

Now sit back, relax, and enjoy!

Press the spacebar to continue.