Photograph or Clipart: Does Object Depiction Affect the Mapping of Language to Referents?

by

Raheleh Haji Gholam Saryazdi

A thesis submitted in conformity with the requirements for the degree of Master of Arts
Psychology
University of Toronto

© Copyright by Raheleh H. G. Saryazdi (2016)
Photograph or Clipart: Does Object Depiction Affect the Mapping of Language to Referents?

Raheleh H. G. Saryazdi

Master of Arts
Psychology
University of Toronto
2016

Abstract

Psycholinguistic studies using the "visual world" paradigm tend to employ either photographs or clipart as stimuli, with convenience or availability typically serving as the sole reason for choosing a particular image type. The present study examined how image type (clipart/photograph) affects the process of mapping language to visual referents. On each trial, participants viewed an array of objects and heard a recorded sentence containing either a semantically neutral verb (Experiment 1: Jamie will move the banana) or a semantically restrictive verb (Experiment 2: Jamie will peel the banana). Image type (clipart/photograph) was manipulated across blocks. Eye movements were recorded as participants listened to the sentences and mouse-clicked on the last-mentioned object. Quite strikingly, image type did not appear to affect language-driven eye movements. However, it did sometimes influence the initial moments of visual processing, before the recorded sentence began. The results are described in terms of their implications for visual-world studies of language processing.
Acknowledgments

I was once told that a great advisor is one who is a great mentor, and I am so grateful to have found that in my advisor Craig Chambers. I would like to thank him for his continuous support, guidance, and mentorship during the course of my thesis. His enthusiasm created an exceptional academic setting which not only fostered my growth as a scholar but also my love for academia. I truly could have not asked for a better advisor. I would also like to thank my committee members, Meredyth Daneman and Elizabeth Johnson, for their valuable suggestions and constructive feedback. I wish to extend my gratitude to Ben Bauer and Nancie Im-Bolter for encouraging me to go to graduate school and for continuing their support through my journey as a graduate student.

I am grateful to the Social Sciences and Humanities Research Council of Canada for their financial contribution to my Master’s thesis. The timely completion of this project would have not been possible without the hard work and dedication of the PAL lab’s research assistants, Agatha Rodrigues, Chris Klammar, Julie Bannon, and Amit Kamdan. I am so lucky to be a member of such a wonderful lab. I would also like to thank Melissa Paquette-Smith for always being available to answer questions and for being a great peer mentor. Also a genuine thanks to Ann Lang, for making the transition to graduate school such a smooth and pleasant process.

I truly consider myself fortunate to be surrounded by such amazing friends, without their support and encouragement I may never have survived my first year of graduate school. Special thanks to Cassandra Aked, Anna Harrington, and Chris Letke for always being there when I needed a friend or when I just needed a study buddy. I would also like to thank all my new friends, my fellow graduate students, who started this journey with me. I am so glad to be part of a cohort who is so supportive of each other.

A very sincere thanks to my mother, who always believed in me, and without whom I would have never made it this far. It is not easy being a graduate student and a parent and I have received tremendous support from everyone for which I am truly grateful for, but most importantly I am thankful for the two understanding angels in my life. To my children Klarissa and Kaiden, you are the reason Mommy has made it this far. This is for you.
# Table of Contents

Acknowledgments ........................................................................................................................................ iii

Table of Contents ....................................................................................................................................... iv

List of Tables ................................................................................................................................................ vi

List of Figures ............................................................................................................................................... vii

List of Appendices ....................................................................................................................................... viii

Chapter 1 Background and Overview ......................................................................................................... 1

1 Introduction ............................................................................................................................................... 1

1.1 Perceptually-Driven Effects in the Visual World Paradigm ................................................................. 3

1.1.1 Shape .............................................................................................................................................. 4

1.1.2 Colour ............................................................................................................................................ 6

1.2 Does the Iconicity of an Image Matter? ................................................................................................. 7

1.2.1 Iconicity in development .............................................................................................................. 8

1.2.2 Iconicity in adults ........................................................................................................................ 10

Chapter 2 Does Iconicity Affect Incremental Referential Mapping? ......................................................... 12

2 Experiment 1 .......................................................................................................................................... 13

2.1 Method ............................................................................................................................................... 13

2.1.1 Participants .................................................................................................................................... 13

2.1.2 Materials ....................................................................................................................................... 13

2.1.3 Procedure ..................................................................................................................................... 18

2.1.4 Data analysis ............................................................................................................................... 18

2.2 Results and Discussion ....................................................................................................................... 19

2.2.1 Prior to sentence onset ................................................................................................................. 19

2.2.2 Noun region .................................................................................................................................. 21

2.2.3 Mouse click latency ....................................................................................................................... 23

Chapter 3 Does Iconicity Affect Anticipatory Eye Movements? ............................................................. 26

3 Experiment 2 .......................................................................................................................................... 29

3.1 Method ............................................................................................................................................... 29
3.1.1 Participants .................................................................................................................. 29
3.1.2 Materials and procedure ............................................................................................ 29
3.1.3 Data analysis ................................................................................................................ 29

3.2 Results and Discussion .................................................................................................... 30
3.2.1 Prior to sentence onset ............................................................................................... 30
3.2.2 Verb region .................................................................................................................. 31

Chapter 4 General Discussion ............................................................................................. 34

4 Does the Iconicity of an Image Matter? .............................................................. 34

4.1 Methodological Considerations and Future Directions ......................................... 36
4.1.1 Stimuli ......................................................................................................................... 37
4.1.2 Preview duration ......................................................................................................... 38
4.1.3 Population .................................................................................................................. 40

4.2 Conclusion ...................................................................................................................... 41

References .......................................................................................................................... 42

Appendices .......................................................................................................................... 51
List of Tables

Table 2.1. List of Objects Presented in the Critical Trials of Experiments 1 and 2 ......................... 17

Table 2.2. Main and Interaction Effects as a Function of Image Type in Experiment 1 ............... 23

Table 3.1. Main and Interaction Effects as a Function of Image Type in Experiment 2 ............... 33
List of Figures

Figure 1.1. Example of visual display in a typical visual world experiment .................................. 3

Figure 1.2. Example display adapted from Dahan & Tanenhaus (2005) ........................................ 5

Figure 2.1. Sample array for (a) photograph condition and (b) clipart condition. Target referent is wine, alternative target is beer, and ruler and duck are unrelated objects ........................................ 15

Figure 2.2. The effect of image type prior to sentence onset for the (a) number of fixated objects, and (b) average fixation duration in Experiment 1 ........................................................................ 20

Figure 2.3. Proportion of fixations to target as a function of image type ......................................... 22

Figure 2.4. The effect of lexical frequency on reaction time (ms) .................................................. 24

Figure 3.1. The effect of image type prior to sentence onset for the (a) number of fixated objects, and (b) average fixation duration in Experiment 2 ........................................................................ 31

Figure 3.2. Proportion of fixations to the target and alternative target as a function of image type and verb type in Experiments 1 and 2 ........................................................................ 32
List of Appendices

Appendix A. Matched Photograph and Clipart Image Pairs of Critical Targets..........................51

Appendix B. Norming Experiments............................................................................................54

Appendix C. LSA Pairwise Cosine and SUBTLEXus Frequency Measures.................................56
Chapter 1
Background and Overview

Human beings appear to have a natural ability to process and comprehend spoken language, but this seemingly effortless process involves intricate underlying mechanisms. For example, as a spoken sentence unfolds, the incoming auditory information (linguistic input) is mapped onto lexical-conceptual representations (associated meanings), which in turn are mapped to specific actions and entities in the relevant context (Dahan & Tanenhaus, 2004; Knoeferle & Guerra, 2012; Kutas & Federmeier, 2000; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). The current study examines the mapping of language to objects in the immediate visual environment, and in particular the influence of the kind of visual objects to which language is mapped. The study focuses on the visual world paradigm, a common experimental methodology for studying real time spoken language comprehension. In this paradigm, eye movements are recorded as participants listen to spoken language while simultaneously viewing a visual array of objects or scenes (Tanenhaus et al., 1995). Despite the important role of the visual environment within this experimental paradigm, little attention has been given to the kinds of objects that serve as referential stimuli across studies. Therefore, the aim of the present study is to examine if and to what extent the manner of pictorial depiction (photograph or clipart) for referential objects affects the course and nature of on-line language processing.

1 Introduction

The majority of the published literature in experimental psychology relevant to object perception has used paradigms that involve 2-dimensional (2D) images (Snow, Skiba, Coleman, & Berryhill, 2014). Studies have varied extensively both in terms of the image type used and the presentation format. Objects have been depicted as photographs, line drawings, or clipart; further, they have been presented either individually, in an array, or in the context of a simple or a more complex scene. In many cases the actual reason behind choosing a more or less realistic image format is not discussed, and it often appears that this potentially important methodological decision is merely a reflection of the experimenter’s personal preference or what was convenient at the time. To a large degree, this state-of-affairs also seems to hold for visual world studies, where line drawings, clipart images, and photographs of objects presented in arrays or scenes
have all been used, in addition to real objects (e.g., Allopenna, Magnuson, & Tanenhaus, 1998; Altmann & Kamide, 1999; Staub, Abbott, & Bogartz, 2012; Tanenhaus et al., 1995)

Although the influence of image type has not been systematically explored in the context of real-time language processing experiments, a more general concern about this issue has been around for decades. In 1976, a study by Bartram which involved matching pictures to real objects in the immediate environment revealed a significant difference in performance when stimuli were line drawings (e.g., reaction times for identical-view images were faster than for different-view images), whereas no differences were observed with photographs. In his conclusion, he warns against the similar treatment of image types, “One point which clearly emerges from the present study is the danger of regarding different modes of pictorial representations as equivalent” (p. 602). In the following decade, one of the most commonly implemented stimulus sets in experimental psychology was a set of black and white line drawings from Snodgrass and Vanderwart (1980). Notably, an updated version by Rossion and Pourtois (2004) that added texture, surface detail, and colour to the original line drawings revealed that these features, particularly colour, significantly improved object recognition.

The relevance of colour in particular is not surprising because a substantial amount of visual processing is devoted to colour information (Bramão, Reiss, Petersson, & Faísca, 2011; Tanaka, Weiskopf, & Williams, 2001). Interestingly, however, the value of colour information for object recognition appears to vary. Some objects are associated strongly with a specific colour (e.g., strawberries are red) whereas others can take on various colour properties (e.g., gloves can be any colour). Therefore, colour diagnosticity — the degree to which an object is associated with a particular colour — could play an important role in real world object representation (Bramão et al., 2011; Tanaka et al., 2001). Indeed, a meta-analysis conducted by Bramão and colleagues (2011) found that the influence of colour was greater for colour diagnostic images than nondiagnostic images, suggesting that object recognition involves access to a conceptual representation for a prototypical object. Furthermore, colour diagnosticity appears to correlate with object category. Items that are manufactured artefacts (e.g., furniture) typically do not depend on colour semantic information to the same extent as living things (e.g., fruits, vegetables, animals). However, even within subcategories of objects, the extent to which colour influences object representation can be further influenced by the task and the identity of the
objects in question. For example, the ability to distinguish lemon from lime is dependent more on colour than distinguishing lemon from pineapple (Bramão et al., 2011; Gale, Laws, & Foley, 2006; Rossion & Pourtois, 2004; Tanaka et al., 2001).

However, evidence that various kinds of perceptual detail play a role recognizing and differentiating visual objects does not entail that this information plays a strong role in all experimental tasks and paradigms, especially those where object recognition is just one component of the task faced by participants. The following section considers the extent to which perceptual-level factors matter when language is incrementally mapped to visual objects that have already been viewed by a perceiver.

1.1 Perceptually-Driven Effects in the Visual World Paradigm

As described earlier, in the visual world paradigm, participants listen to spoken language while simultaneously viewing a display of objects or scenes. One of the critical findings from studies employing this paradigm is that as language unfolds, listeners’ eye gaze moves rapidly to a denoted target. Thus, if participants hear the word cake in an unfolding sentence while simultaneously viewing an array of objects like the one depicted in Figure 1.1, they will rapidly map the incoming speech to the intended object on the display as the sounds in the word cake unfold in time.

*Figure 1.1. Example of visual display in a typical visual world experiment*
Because the display is typically presented before the language starts, one possibility is that the visual information is initially recoded as an internal mental representation that simply associates concepts with locations. For example, given a visual display of four objects like the one illustrated in Figure 1.1, the display might be encoded into a list-like representation of the form: CAKE - top left, DOLL - top right, TAPE - bottom left, SUNGLASSES - bottom right. The incoming linguistic input would draw on this mental representation to drive eye movements; thus upon hearing cake, participants will look towards the relevant object because they are relying on this type of stored association (CAKE - top left). On this account, the actual visual percept plays a minor role at the point where the noun is heard. However, the evidence suggests this is not, in fact, the correct explanation. A number of studies have shown that perceptual information is not only taken up but is also influential in directing gaze as speech unfolds in real time. In particular, salient properties of the stimuli, such as shape and colour, have shown to significantly determine eye movement behaviour during the course of on-line language processing.

1.1.1 Shape

One important feature of the visual world paradigm concerns listeners' tendency to momentarily consider displayed objects other than the target. These patterns are not random, but often reflect specific kinds of relationships between objects receiving momentary visual consideration and the target object that has been named. One effect found in a number of studies is that, as speech unfolds, listeners are more likely to direct eye movements towards objects whose shape properties are similar to those of a denoted target (e.g., Dahan & Tanenhaus, 2005; Huettig & Altmann, 2007; Rommers, Meyer, & Huettig, 2015; Rommers, Meyer, Praamstra, & Huettig, 2013; Yee, Huffstetler, & Thompson-Schill, 2011). An example of this is provided in Figure 1.2. When presented with a display of this sort, participants’ gaze patterns show an elevated likelihood of briefly fixating the image of a rope upon hearing snake, before the actual snake is fixated. This occurs despite the absence of any other conceptual relationship (Dahan & Tanenhaus, 2005; see also Huettig & Altmann, 2007). Importantly, this pattern would be unexpected on an account whereby objects had simply been identified earlier in terms of their conceptual category and stored in a list like format with their associated spatial locations in the visual array. That is, if SNAKE was paired with "top right", there is no reason to expect listeners to make an "accidental" fixation to the bottom right any more than an accidental fixation.
elsewhere in the display. The observed pattern instead suggests that perceptual features of objects are important at the precise time where the noun is mapped to referential candidates: objects sharing the same perceptual features can interfere with this process to some degree.

**Figure 1.2.** Example display adapted from Dahan & Tanenhaus (2005).

In subsequent work, Yee and colleagues (2011) explored whether this type of perceptual competition effect depends only on the actual features of the displayed objects or also involves listeners' abstract knowledge about *canonical* object shape. In the "shape" condition, the target was an object that prototypically has the same shape characteristics as a competitor object, but was not depicted in that form. For example, the target word *pizza* was depicted as a slice (triangular shape). The competitor was a Frisbee (circular shape), which shared the same shape features as a prototypical representation of a pizza (i.e., round). In a second condition, target-competitor pairs were related in terms of their function (e.g., tape and glue). Although there was a greater degree of semantic overlap between items in the function condition, the consideration of the competitor upon hearing the target name was observed only in the shape condition, clearly illustrating the stronger influence of perceptually driven competition, even when the relevant perceptual features are not instantiated in the target object (Yee et al., 2011; Experiment 1).

Similar findings have also been reported by Rommers and colleagues (2013, 2015) who presented their participants with three different trials including the target, the shape competitor, or a control item alongside three unrelated items. Upon hearing a sentence that began with “*In 1969, Neil Armstrong...*”, participants fixated more on an item that shared a similar shape as an
associated concept such as MOON (e.g., tomato) than an item whose shape characteristics bore no such relationship (e.g., rice). Here, the anticipation of the word MOON elicits the activation of the concept of ROUND and thus drives fixations to round objects such as a tomato despite the absence of semantic association between MOON and TOMATO (Rommers et al., 2013, 2015).

1.1.2 Colour

A series of experiments by Huettig and Altmann (2004, 2011) examined how colour drives shifts in visual attention during language processing. The authors found that when participants were given an array of four items, one of which (e.g., lettuce) was related to the spoken noun (e.g., frog) in terms of a shared colour feature (green), the likelihood that eye movements were temporarily directed towards the colour competitor (e.g., lettuce) was greater than for unrelated items (e.g., mitten, pipe, and suitcase; Huettig & Altmann, 2004). More recently, Huettig and Altmann (2011) examined whether words whose referent was not actually depicted (e.g., pea) triggered the visual consideration of an item bearing the relevant colour (e.g., green blouse, for which green is not diagnostic, unlike lettuce) versus an item that was conceptually related (e.g., mushroom, both are vegetables). The authors found elevated fixations to these objects in both cases and thus concluded that real-time language processing is rapidly influenced by both perceptual and conceptual information of this sort.

Similar perceptually driven effects have also been demonstrated in studies that examine eye movement behaviour in children. In a study by Johnson and Huettig (2011), 36-month-olds were more likely to look at an object on the display (e.g., red plane) that was associated with the spoken word (e.g., strawberry) in terms of its shared colour attribute (red) than an unrelated object (e.g., yellow plane). In a follow-up study, Johnson, McQueen, and Huettig (2011) found similar effects with an even younger group of 24-month-old children who were not yet versed in their colour vocabulary. These toddlers would look at yellow objects upon hearing banana but not when they explicitly heard the colour yellow, which suggests that perceptually driven effects of this sort are not mediated by linguistic-conceptual knowledge (e.g., some type of spreading activation from the lexical entry BANANA to YELLOW to the yellow airplane). This is because YELLOW was not yet part of the child's vocabulary. In sum then, findings from both adult and children literature clearly illustrate that real time language comprehension is guided by colour information available in the immediate visual environment. Again, effects of this sort would be
unexpected if eye movements were driven solely by conceptual category information that had been extracted when the visual display was originally viewed, before the linguistic stimuli were heard.

1.2 Does the Iconicity of an Image Matter?

Although photographs and clipart images that depict the same concrete object share similar perceptual features at a coarse level (e.g., orientation, shape features, colouring), they differ in important ways. Line drawings lack surface detail and thus are a more simplistic depiction of a real object in comparison to photographs, whereas photographs’ detailed features make them more iconic or faithful as a representation of the original object (Bartram, 1976; Brodie, Wallace, & Sharrat, 1991; Lawson & Humphreys, 1996; Oates & Reder, 2011; Salmon, Matheson, & McMullen, 2014; Wimmer, Robinson, Koenig, & Corder, 2013). The presence of realistic colours and shadows in photographs also contributes strongly to their greater level of iconicity (the extent to which a depicted image represents a real object; Pierroutsakos & DeLoache, 2003). Even black and white photographs have more depth detail than line drawings, and are often considered as more representative of the real objects (Gagnier & Intraub, 2012; Huettig & Altmann, 2011).

The less-realistic nature of some of the items used in visual world studies (i.e., line drawings, clipart) has only been recently discussed and to date only a handful of studies have assessed whether comparable effects are found with realistic scenes (e.g., Andersson, Ferreira, & Henderson, 2011; Staub et al., 2012). For example, Staub and colleagues (2012) replicated an earlier clipart-based study by Altmann and Kamide (1999) in order to determine if the same language-driven gaze patterns are observed with photographic images, where the depictions are more naturalistic. Similarly, Andersson and colleagues (2011) examined whether realistic scenes (even more complex and "cluttered" than the ones used by Staub et al., 2011) impact the way in which gaze patterns reflect the incremental interpretation of unfolding language. Both studies showed effects that in fact appeared to be comparable to those in studies using clipart imagery. However, these conclusions were based on the overall similarity of the patterns with those reported in the previous literature and did not involve a direct comparison of photographs versus clipart within the same experimental context and task. It is therefore an open question whether the iconicity of an image would influence the nature and course of real-time language processing.
at a more fine-grained level. Further, if there is a difference between more and less iconic images, in what direction would the difference lie? For example, would more iconic stimuli facilitate real-time language comprehension due to their naturalness, or would they be more taxing because of a greater demand placed on cognitive resources to maintain or process additional visual detail? It is also possible that processing differences that arise might not be related to complexity *per se*, but could instead reflect the fact that various image types are conceptualized differently. In fact, there is evidence from developmental research suggesting that this may be the case.

### 1.2.1 Iconicity in development

Although 2D stimuli in lab experiments are clearly different than the real objects we interact with in daily life, our experience with such media begins early. Children learn many concepts that they have never seen (e.g., tugboat) or will never see (e.g., dinosaurs, fairies) from picture books (Troseth, Pierroutsakos, & DeLoache, 2004). Interestingly, pictures of concrete objects, whether depicted as a photograph, line drawing, or clipart, are understood to play a dual representational role in the mind of the perceiver, as they can be conceptualized in terms of what they are (in terms of its actual 2D physical characteristics, e.g., ink on a flat piece of paper) but also for the 3D object they “represent” in broad terms (DeLoache, 2004; DeLoache, Pierroutsakos, & Uttal, 2003; Gelman, Waxman, & Kleinberg, 2008; Gibson, 1978; Pierroutsakos & DeLoache, 2003; Preissler & Bloom, 2007; Simcock & DeLoache, 2006; Troseth et al., 2004). In fact, pictorial competence — the ability to see beyond the physical surface of the 2D image and understand its symbolic representation — is not innate, but learned (DeLoache et al., 2003; Simcock & DeLoache, 2006; Troseth et al., 2004). Previous studies suggest that children do not develop pictorial competence until the age of 2, but are able to discriminate a picture from a real object. For example, infants and toddlers show a preference for real objects than images when presented with both; thus, they can discriminate between 2- and 3-dimensional objects even though they lack an understanding of their relationship (Pierroutsakos & DeLoache, 2003; Troseth et al., 2004). Further, the degree to which children manually explore images (e.g., physically interact with the image as if it was the real object itself) depends on the level of iconicity. Pierroutsakos and DeLoache (2003) presented 9-month-olds with four different types of stimuli (i.e., colour photograph, black-and-white photograph, colour line
drawing, and black-and-white line drawing) and observed how the infants interacted with the images. The greater the overlap between the depicted and the real object it represented (i.e., colour photograph), the greater the extent of manual exploration by the infants (Pierroutsakos & DeLoache, 2003, Experiment 1).

The topic of iconicity has also been explored in studies examining the role of pictures in the acquisition of language and general knowledge (Gibson, 1978; Khu, Graham, & Ganea, 2014; Simcock & DeLoache, 2006; Tare, Chiong, Ganea, & DeLoache, 2010; Troseth et al., 2004). One topic relevant here is whether iconicity or pictorial realism affects learning outcomes. This research has found that picture books containing images with a high degree of iconicity (i.e., photographs) improve general learning more than those with less iconic images in young children (i.e., line drawings; Simcock & DeLoache; Tare et al., 2010).

Interestingly, a study by Wimmer and colleagues (2013) conducted with preschoolers (3- to 5-year-olds) revealed that even older children are sensitive towards the level of image iconicity. However, these effects were mediated by the type of task. The authors first examined children’s performance on matching real world objects to one of three different image depictions (i.e., photograph, colour, and black-and-white line drawing). For example, they first presented children with a doll, as well as a picture of the same doll, both wearing a sticker. The experimenter then either changed the sticker on the doll (e.g., from boat to butterfly), while the image was turned over, or vice versa. The results revealed that regardless of the image type, children (erroneously) expected the change to be transferred over, thus the greater degree of realism of a photograph did not eliminate what they referred to as “representation-referent confusion” (incorrect transfer of attributes from 3D to 2D object or vice versa). However, in a follow-up experiment involving a simple recognition task, there was evidence that children are, in fact, influenced by the degree of image iconicity. For example, when children were presented with a real object that included an additional feature (e.g., postman doll with a bag) and then were asked which of the three different image representations "best matched" the given object (where only one image actually included the bag), they would preferentially select photographs over line drawings. The photograph was thus recognized as a more realistic representation of the 3D object, even when it was not an exact match (e.g., no bag for the postman doll). Thus, these findings suggest that children are sensitive towards differences in the iconicity of images in
certain tasks. The relevance of these developmental effects for adults rests in part of earlier findings showing that effects observed in children’s overt behaviour is often reflected in adult’s implicit and momentary processing behaviour (e.g., so-called egocentrism effects: Keysar, Barr, Balin, & Brauner, 2000).

1.2.2 Iconicity in adults

Only a handful of studies conducted with adults have explored the effect of image iconicity, and the focus of this work has been object recognition. In a study by Salmon and colleagues (2014), objects that were manipulable (i.e., that could be picked up and interacted with) were better recognized when presented as photographs than as line drawings, thus the surface feature information seemed to significantly influence recognition. The basis of the observed effect was attributed to facilitation from activated motor representations in the manipulable photograph condition due to the more realistic depiction of an object (Salmon et al., 2014). A study by Sareen, Ehinger, and Wolfe (2015) examined whether objects presented in the non-reflected part of a photograph of a scene containing a mirror (e.g., bathroom scene with some objects reflected in the mirror) were perceived differently than the same objects presented in the mirrored reflection. In their first experiment, participants were asked to label the objects on the photograph, whereas in the second experiment they were asked to detect changes between a new photograph and a previously presented photograph. Participants were more likely to label and detect changes in the non-reflected part of the image than the reflected part. Participants viewed the depicted mirror in a photograph as having the same properties of a real mirror, hence, reflections in the mirror were perceived as "less real" and therefore were explored with less vigilance than the other objects in the depicted scene.

These aforementioned studies clearly illustrate various kinds of processing benefits for more realistic images (and even images that are simply perceived as more realistic). There are, however, contrasting results. Snow and colleagues (2014) examined recall and recognition performance in a study comparing real objects, coloured photographs, and black and white line drawings. Although recall and recognition for real objects was better than photographs and line drawings, there were no differences observed between the two image types. In fact, the authors concluded that the photographs’ more realistic colouring and texture did not make any additional contribution. The extent to which these findings could be generalized to the visual world
paradigm, however, is not yet known. The next chapter will describe a visual world experiment designed to systematically examine the effect of image type on real-time language processing.
Chapter 2
Does Iconicity Affect Incremental Referential Mapping?

In the visual world paradigm, the measures of interest are based on linguistically-mediated eye movements to objects in the concurrent visual display. To date, no study has considered the extent to which the process of mapping speech to referents is influenced by the type or quality of images presented within the context of the same experiment. The use of clipart and similar kinds of depictions is not only relevant for questions about the ecological validity of studies of real time language comprehension (Andersson et al., 2011; Staub et al., 2012) but also for questions about the precise nature of the referential processes that occur in these experimental contexts. Consider, for example, that an utterance like *Click on the snake* in fact serves as a kind of linguistic shortcut for saying *Click on the picture of the snake*, or something similar, seeing as the referent is not a live reptile. Thus, listeners' ability to quickly map incoming speech to a visual referent in this context reflects the tendency for people to behave as "cooperative" language users (Grice, 1975). Listeners understand the reference even though the noun phrase that was heard is, in a technical sense, incorrect. The need to cooperate in this way when understanding referential descriptions is widespread, such as when a person hears *I'm parked out back* (rather than *My car is parked out back*) or *This is my favourite author* as the speaker holds a novel in the air. One possibility is that the processing that underlies this cooperative behaviour differs for images that vary in their iconicity. Thus, the goal of the current study is to offer insight into the extent to which pictorial depiction of concrete objects affects the mapping of language to referents. This will be accomplished by measuring the degree to which fixation proportions and saccadic latencies to target objects differ as a function of the use of photographs or clipart as stimulus images.

In the present study, the implementation of clipart and photograph stimuli differs from the stimulus set used in any previous visually situated language comprehension study. Every 2D target item in the current experiment began as a real 3D object purchased, photographed, and then digitally transformed to create a corresponding clipart version for the photograph. Also, the display on each trial consisted of a visual array of objects rather than a naturalistic visual scene in order to examine the effects of interest independent of the various kinds of additional contextual information involved in naturalistic scenes (Henderson & Hollingworth, 1999). It is
plausible that photographs, with their high iconicity, might facilitate the mapping of the linguistic information to its referent in comparison to clipart images, which are considered to be less iconic. Alternatively, clipart images, with their vibrant colours and clear contours, or their more “abstract” meanings may stand out more and thus streamline the mapping process. A third possibility is that the type of object depiction may not have any influence on the mapping process.

The potential implications of the present study are twofold. In terms of methodological implications, the presence of any significant difference as a function of stimulus type would urge researchers to acknowledge the role of pictorial representation prior to comparing and challenging results across studies. Alternatively, in the absence of any observed differences, this will be the first study to reveal that the methodological variation in terms of visual stimuli do not influence on-line language processing and thus researchers could retain their preferred form of presentation. A second implication of the study is theoretical, as any evidence of differences in the process of mapping of language to photographs versus clipart will suggest that different mechanisms or representations are at play.

2 Experiment 1

2.1 Method

2.1.1 Participants

Forty students (Age range = 17 to 26, M = 19.23, SD = 1.89) were recruited from the University of Toronto Mississauga to participate in the current study. They received a course credit or $10 per hour as compensation. Participant selection was limited to individuals with native-like proficiency in English, normal hearing, and normal or corrected to normal vision.

2.1.2 Materials

Visual stimuli. The stimulus set for critical trials comprised of 96 distinct images of concrete objects (24 target images, 24 alternative target image, and 48 other images), four of which would be presented together on a given trial. These were presented either as photographs or as clipart objects (meaning that a total of 192 images were prepared). The items were selected from various object categories (i.e., household objects, clothing, food) in order to ensure that they
were representative of the kind of real world objects individuals interact with on a daily basis (see Table 2.1). Care was taken to ensure none of the four objects occurring on a single display were perceptually similar, as there is often an attentional bias towards objects that share similar perceptual properties (e.g., Huettig & Altmann, 2007), particularly if the objects are familiar (Rouder, Ratcliff, & McKoon, 2000). As noted above, the images for critical trials were created especially for the experiment from photographs of real 3D objects. The objects were photographed with a Nikon D90 DSLR camera fitted with a Nikon Micro-Nikkor 40mm f/2.8G zoom lens. When possible, photographs were taken using the optimal-centered view, which is considered to facilitate the mapping of object images to existing mental representations (Lawson & Humphreys, 1996). Three raters assessed the quality of each photographic image prior to their digital transformation to clipart. Photographs of target items were selected based on 100% agreement (otherwise photographs were retaken). Digital editing software (Gimp version 2.0) was used to create a clipart image corresponding to each photograph. The clipart images matched the colours of the original photographs but the colours were more uniform and brighter (and thus more perceptually salient), consistent with the general nature of clipart imagery. The clipart versions also had a reduced amount of surface detail (texture, shading) and had defined black outlines (see Appendix A). To ensure that the critical images were clearly identifiable by participants, a separate set of recognition norming experiments were also conducted prior to final image selection (see Appendix B).

A remaining set of 96 images (48 clipart and 48 photographs) was assembled for use on filler trials. These images were downloaded from an online commercial image repository, and care was taken to select photograph and clipart versions of each object that were similar in look and angle of perspective. In instances when it was not possible to obtain a comparable photograph and clipart for a given object, a real 3D object was photographed and a corresponding clipart version was created, following the same procedure used for critical trials. All images from critical and filler trials were edited to match in brightness and contrast in order to ensure consistency in overall visual presentation. If necessary, images were resized to ensure size representativeness in relation to other items on a particular trial (e.g., resized toothbrush to be smaller relative to scarf). The images were displayed on a 22-inch LCD monitor at 375 x 375 pixel resolution in a 2 x 2 grid format. See Figure 2.1 for a sample display of an object array in the clipart and photo conditions.
Figure 2.1. Sample array for (a) photograph condition and (b) clipart condition. Target referent is wine, alternative target is beer, and ruler and duck are unrelated objects.
**Auditory stimuli.** To allow comparisons with a follow-up study examining the effect of the verb information on referential processing, the present study implemented a variation of “look and listen” paradigm where participants heard the target object within a declarative sentence with a subject identified by a unisex name (i.e., Jamie). Thus all sentences were in the form of *Jamie will move the apple*. Listeners were instructed to use a mouse to click on the object that “Jamie” will move. (As will be seen in Experiment 2, materials of this sort allow the informativity of the verb information to be easily manipulated, e.g., *Jamie will peel the apple.*) Sentences were recorded in neutral intonation by a female native English speaker and were digitally edited to have an average intensity level of 70 dB SPL.

The precise target for each critical trial was varied across different sub-versions of the experiment. This was done to ensure that any possible effects of image type (photograph vs. clipart) would be unlikely to result from the particular object chosen as a target. Thus, two objects in each display could serve as the target mentioned in the corresponding sentence, alternating across lists (e.g., *Jamie will move the beer, or Jamie will move the wine*). Given the potential influence of lexical frequency on language processing, these lists were balanced for target name frequency using SUBTLEXus, which measures the frequency of words that appear in a large-scale corpus built from movie subtitles, and thus is more representative of spoken language than frequency measures based on written language (see Appendix C). Specifically, the stimulus set was overall balanced in terms of the pairing of high frequency (HF) and low frequency (LF) nouns as targets on critical trials. See Table 2.1 for the list of objects used on critical trials.
Table 2.1. List of Objects Presented in the Critical Trials of Experiments 1 and 2

<table>
<thead>
<tr>
<th>Category</th>
<th>Target HF</th>
<th>Target LF</th>
<th>Unrelated 1</th>
<th>Unrelated 2</th>
<th>Expt. 2 Verb</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clothing</td>
<td>Hat</td>
<td>Sunglasses</td>
<td>Tongs</td>
<td>Lightbulb</td>
<td>Wear</td>
</tr>
<tr>
<td>Clothing</td>
<td>Tie</td>
<td>Socks</td>
<td>Paintbrush</td>
<td>Knife</td>
<td>Fold</td>
</tr>
<tr>
<td>Clothing</td>
<td>Shirt</td>
<td>Scarf</td>
<td>Dart</td>
<td>Toothbrush</td>
<td>Wash</td>
</tr>
<tr>
<td>Clothing</td>
<td>Bag</td>
<td>Umbrella</td>
<td>Sponge</td>
<td>Paddle</td>
<td>Open</td>
</tr>
<tr>
<td>Clothing</td>
<td>Shoe</td>
<td>Belt</td>
<td>CD</td>
<td>Pear</td>
<td>Tighten</td>
</tr>
<tr>
<td>Household</td>
<td>Safe</td>
<td>Chest</td>
<td>Toilet Paper</td>
<td>Drumstick</td>
<td>Close</td>
</tr>
<tr>
<td>Household</td>
<td>Lamp</td>
<td>Flashlight</td>
<td>Bottle</td>
<td>Nail polish</td>
<td>Turn on</td>
</tr>
<tr>
<td>Household</td>
<td>Candle</td>
<td>Lantern</td>
<td>Rolling Pin</td>
<td>Scissors</td>
<td>Light</td>
</tr>
<tr>
<td>Household</td>
<td>Jar</td>
<td>Mug</td>
<td>Blocks</td>
<td>Nail Clipper</td>
<td>Empty</td>
</tr>
<tr>
<td>Household</td>
<td>Pot</td>
<td>Spoon</td>
<td>Rubik's Cube</td>
<td>Notebook</td>
<td>Clean</td>
</tr>
<tr>
<td>Household</td>
<td>Vase</td>
<td>Pail</td>
<td>Crayons</td>
<td>Hammer</td>
<td>Fill</td>
</tr>
<tr>
<td>Household</td>
<td>Rope</td>
<td>Bow</td>
<td>Fork</td>
<td>Stapler</td>
<td>Untie</td>
</tr>
<tr>
<td>Food</td>
<td>Grapes</td>
<td>Strawberries</td>
<td>Wrench</td>
<td>Tape</td>
<td>Rinse</td>
</tr>
<tr>
<td>Food</td>
<td>Pie</td>
<td>Cookies</td>
<td>Syringe</td>
<td>Envelope</td>
<td>Eat</td>
</tr>
<tr>
<td>Food</td>
<td>Chocolate</td>
<td>Butter</td>
<td>Hanger</td>
<td>Spade</td>
<td>Melt</td>
</tr>
<tr>
<td>Food</td>
<td>Beer</td>
<td>Wine</td>
<td>Toy duck</td>
<td>Ruler</td>
<td>Drink</td>
</tr>
<tr>
<td>Food</td>
<td>Onion</td>
<td>Carrot</td>
<td>Stamp</td>
<td>Bolt</td>
<td>Chop</td>
</tr>
<tr>
<td>Food</td>
<td>Eggs</td>
<td>Noodles</td>
<td>Screwdriver</td>
<td>Clipboard</td>
<td>Boil</td>
</tr>
<tr>
<td>Food</td>
<td>Bread</td>
<td>Lettuce</td>
<td>Marker</td>
<td>Dropper</td>
<td>Cut</td>
</tr>
<tr>
<td>Food</td>
<td>Apple</td>
<td>Banana</td>
<td>Cigarette</td>
<td>Earrings</td>
<td>Peel</td>
</tr>
<tr>
<td>Food</td>
<td>Cake</td>
<td>Cheese</td>
<td>Magnet</td>
<td>Lighter</td>
<td>Slice</td>
</tr>
<tr>
<td>Other</td>
<td>Ball</td>
<td>Dice</td>
<td>Comb</td>
<td>Ladle</td>
<td>Roll</td>
</tr>
<tr>
<td>Other</td>
<td>Recorder</td>
<td>Harmonica</td>
<td>Dustpan</td>
<td>Bowl</td>
<td>Play</td>
</tr>
<tr>
<td>Other</td>
<td>Present</td>
<td>Candy</td>
<td>Soap Dispenser</td>
<td>Lipstick</td>
<td>Unwrap</td>
</tr>
</tbody>
</table>

Note. The sentences in Experiment 1 included the same verb (“move”), whereas the verbs in Experiment 2 were all unique action verbs.
2.1.3 Procedure

Participants were seated approximately 25 inches away from a 22-inch display monitor. Gaze was tracked using an Eyelink 1000 eye-tracker with remote optics (SR Research, Ontario, Canada), positioned under the monitor. Eye position was sampled at 500 Hz, and the system achieves \(0.5^\circ\) average accuracy in visual angle. In the present study, photographs and clipart images were presented in separate blocks and never intermixed on a single trial. It has been shown that visual saliency facilitates attention towards a displayed object prior to an utterance (Coco & Keller, 2015). Thus, a mixed condition of clipart and photographs could have risked boosting attention to clipart due to their brighter and more uniform colouring, making it difficult to parse out the effects of interest.

As noted earlier, each trial included four objects that were displayed on a 2 x 2 grid: the target object and three objects that were not mentioned. The object placements were randomized to ensure the same likelihood of presentation on each corner of display. Further, there was a separate block of photographs and clipart images for each of the four lists, which were balanced for the order of image presentation (photograph first vs. clipart first) and frequency of critical target (HF vs. LF target nouns). The pairing of items to conditions was achieved using a list design. Each participant saw an object array in only one condition (photograph vs. clipart) and with only either HF or LF version of the target object. However, across participants, all item arrays occurred in each condition.

Participants were instructed to listen to sentences and select the corresponding referent on the display using a mouse. The experiment began with two practice trials to ensure understanding of the task. Each trial began with a fixation point in the center of the screen so that participants’ gaze started at a central location, and also to allow for drift correction by the eye tracker. The visual stimuli were presented 3 s prior to the beginning of the accompanying recorded sentence. Experimenter Builder software (SR Research, Ontario, Canada) was used to program the entire experiment. Participants were tested in a single session that lasted approximately 30 minutes.

2.1.4 Data analysis

All statistical analyses were conducted using the open source statistical software R version 3.2.4 (R Core Team, 2015). The eye movement data were analyzed using linear mixed effect models
(lme4 package version 1.1-11; statistical significance was assessed with lmerTest version 2.0-30, Bates, Maechler, Bolker, & Walker, 2015). The linear mixed effect approach allows for fixed effects to be evaluated while simultaneously accounting for the possible variability related to participants and items. Image type (photograph vs. clipart) was entered as a fixed effect, whereas participant and item were entered as random effects with the model including both intercept terms and slopes for the fixed effects. For the analyses conducted prior to sentence onset, an "item" in the random effects structure corresponded to a particular array of objects. For analyses conducted after sentence onset, an item corresponded to the target object that was named.

The data for all regions were coded by quadrant (the four corresponding regions of interest) and in separate 25 ms time bins. All eye movement analyses were based on trials with correct responses (participants selected the incorrect item on only 1% of trials). Latency measures for both eye movements (target saccades) and manual responses (mouse click latency) were inspected for outliers before statistical analysis. Any observation falling outside +/- 2 SD from the mean was replaced by the corresponding lowest or highest value within this distribution. Furthermore, all proportional data (e.g., the relative likelihood of fixating a given display object) were logit transformed prior to statistical analysis.

2.2 Results and Discussion

2.2.1 Prior to sentence onset

One question is whether the type of image has an influence on "early" moments of visual processing after the object array initially appears. Although this consideration is often not part of visual world studies, it is nevertheless important to examine given the current questions of interest. For example, the efficiency of mapping language to referents in the visual environment may be influenced by how easily visual objects were identified. Therefore, any differences that would arise from the processing of images prior to sentence onset may consequently affect the course of linguistic processing after the sentence onset.

Analyses of eye movements during the first 3 s of a trial (when the participants were only presented with the visual input) revealed significant differences in the processing of the two image types. In terms of the number of fixated objects (see Figure 2.2-a), there was a small but significant effect of image type whereby participants inspected more objects in the visual display...
on average when the stimuli were photographs ($M = 3.18$) compared to clipart ($M = 3$), $\beta = .09$, $SE = .04$, $t_{(34.99)} = 2.20$, $p = .034$. Additionally, the average duration of individual fixations (see Figure 2.2-b) was shorter in this time period for photographs ($M = 285.29$) than for clipart images ($M = 318.10$), $\beta = -16.45$, $SE = 7.10$, $t_{(39)} = -2.31$, $p = .026$. A further correlational analysis (across participants) confirmed a significant negative relationship between these two measures (number of fixated objects and individual fixation duration), suggesting the two effects are indeed related, $r_{(948)} = -.63$, $p < .001$. Therefore, the type of pictorial depiction matters for at least the early moments of visual processing. One possible explanation is that participants are more efficient at scanning the visual display with photograph than clipart stimuli due to the more "natural" character of the former. It is, however, also possible that visual features of the clipart stimuli (bold outlines, brighter colouring) captured and held participants’ attention for a longer period. Related to this is the possibility that the observed effects were due to the specific features of the clipart images constructed for this study. To assess this, follow-up analyses were conducted with both critical and filler trials, where the latter consisted primarily of images obtained from the on-line repository. The effect of image type remained significant, and therefore is not related to the custom-created stimulus set.

![Figure 2.2](image)

*Figure 2.2. The effect of image type prior to sentence onset for the (a) number of fixated objects, and (b) average fixation duration in Experiment 1*
2.2.2 Noun region

To assess the incremental mapping of language to display objects, eye movement patterns on critical trials were coded within a 500 ms time window extending from 200 ms to 700 ms after the target noun onset. The selected time interval was chosen to reflect the minimum duration of the target nouns during the critical trials ($M = 711.67$, $Range = 486$ to 1,113 ms), combined with the approximately 200 ms lag required for saccadic eye movements to occur on the basis of incoming speech information. Figure 2.3 plots the proportion of fixations to target objects across the image type condition. Although graphically it seems that participants reached the target image more efficiently with the clipart images as the sound in the target noun unfolded in time, this difference was not significant, $\beta = -.42$, $SE = .29$, $t_{(25.43)} = -1.43$, $p = .164$. Additionally, we conducted follow up analyses controlling for possible noise in the data resulting from differences in lexical frequency and the degree of "descriptive fit" between images and their corresponding labels. Frequency was implemented as a fixed effect in two different ways in two separate models. In one model, frequency was treated as a continuous variable, reflecting the raw SUBTLEXus frequency counts for the target word. In a second model, frequency was treated as categorical (HF vs. LF), reflecting the relative frequency difference between the target and the alternative. The descriptive fit measures were based on the image rating scores in the final Mechanical Turk norming experiment (see Appendix B) and were also entered as a fixed effect.

With the inclusion of the additional fixed effects (i.e., lexical frequency and descriptive fit) into the new models, there was still no significant effect of image type. Furthermore, there was no evidence of any interaction between image type and frequency nor was there an effect of image type by descriptive fit, all $ps > .05$. These results suggest that the type of pictorial depiction is not important for linguistic processing, at least in relation to the mapping of nouns to depicted objects. In other words, the process of mapping the word to a visual referent has the same time course regardless of whether the image is depicted as a photograph or as a clipart.
These null results, however, are not indicative that the present study’s measures were simply insensitive. Figure 2.3 clearly illustrates that the participants’ looks toward the target increased as the noun unfolded in time, which reflects the incremental nature of on-line referential processing. Using a growth curve analyses (GCA) with time (200 to 700 ms intervals) and image type (photograph vs. clipart) as fixed effects and participants as random effects, a main effect of time was observed, confirming the observed incremental pattern. Table 2.2 presents the results of main and interaction effects as a function of image type in Experiment 1.
Table 2.2. Main and Interaction Effects as a Function of Image Type in Experiment 1

<table>
<thead>
<tr>
<th>Effect</th>
<th>Estimate</th>
<th>SE</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before Sound: Number of Fixated Objects</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Intercept)</td>
<td>3.09</td>
<td>.09</td>
<td>32.69</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Image Type</td>
<td>.09</td>
<td>.04</td>
<td>2.20</td>
<td>.034</td>
</tr>
<tr>
<td>Before Sound: Average Fixation Duration (ms)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Intercept)</td>
<td>301.69</td>
<td>15.08</td>
<td>20.01</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Image Type</td>
<td>-16.41</td>
<td>7.10</td>
<td>-2.31</td>
<td>.026</td>
</tr>
<tr>
<td>Noun Region: Proportion of Fixations to Target</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Intercept)</td>
<td>-1.52</td>
<td>.62</td>
<td>-2.45</td>
<td>.018</td>
</tr>
<tr>
<td>Image Type</td>
<td>-.42</td>
<td>.29</td>
<td>-1.43</td>
<td>.164</td>
</tr>
<tr>
<td>Noun Region: Growth Curve Analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Intercept)</td>
<td>-.13</td>
<td>.02</td>
<td>-7.96</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Time</td>
<td>.02</td>
<td>.002</td>
<td>11.11</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Image Type</td>
<td>-.0002</td>
<td>.01</td>
<td>-.02</td>
<td>.983</td>
</tr>
<tr>
<td>Time x Image Type</td>
<td>-.001</td>
<td>.0004</td>
<td>-1.17</td>
<td>.243</td>
</tr>
</tbody>
</table>

2.2.3 Mouse click latency

In addition to eye-movement measures, reaction time information was also obtained during each trial. This measure was based on the time (ms) it took participants to click on the target noun, timed from the noun's onset. Although there was no effect of image type (Photograph $M =$
1567.22 ms, Clipart $M = 1562.66$ ms), the results revealed an overall (collapsed across conditions) significant effect of relative lexical frequency, whereby relatively high frequency target nouns ($M = 1486.04$, $SD = 368.50$) led to faster target mouse-clicks than low frequency target nouns ($M = 1654.70$, $SD = 398.92$), $t_{(38)} = -8.69$, $p < .001$. Figure 2.4 provides an illustration of this effect. Therefore, both the results from the online and offline measures reflect well-established linguistic phenomena (i.e., incrementally, lexical frequency), further confirming that the current study’s measures should in principle be sensitive enough to detect an effect of image type if it had been present.

![Figure 2.4. The effect of lexical frequency on reaction time (ms)](image)

In sum, the results of the current experiment clearly demonstrate that the type of pictorial depiction matters, but only during the initial scanning of the visual display (before any language had been heard). More specifically, the significant effects of image type (a greater number of objects inspected with photographs, co-occurring with shorter individual fixation durations than for clipart) were only observed during the early moments of visual processing, (i.e., during the
phase of visual object recognition). There was no residue of this effect by the time information from the incoming speech stream was received, possibly because by that point the visual percepts were already mapped onto mental representations of their corresponding conceptual categories and thus the process of object recognition was already complete. These results are in accordance with scene based paradigms in visual cognition experiments, where initial viewing behaviour at stimuli onset has been shown to differ from viewing behaviour several seconds into the presentation (Tatler, 2016).

Although the present study’s measures were sensitive enough to capture common psycholinguistic effects (i.e., incrementality, lexical frequency), no effects on referential mapping were found. The current experiment, however, examined the process of mapping nouns to objects, which is only one way to test for effects of image iconicity on language processing. Another approach is to examine the mapping of verbs to objects, which is mediated by an understanding of the actions that certain objects could participate in. This is explored in the following chapter.
Chapter 3
Does Iconicity Affect Anticipatory Eye Movements?

In a seminal paper by Altmann and Kamide (1999) and its recent adaptation by Staub and colleagues (2012), participants’ gaze was found to shift to the image of a cake in a visual display as soon as they heard the verb eat in the sentence The boy will eat the cake (cupcake in Staub et al., 2012). This study was one of the first to illustrate that real-time language processing is not only incremental but also anticipatory in nature. Participants anticipated the target cake from other alternatives (e.g., boy, book, toy) based on the restricted action imposed by the verb eat. This knowledge of what actions an object could afford – that a particular entity (e.g., cake) could evoke a particular action (e.g. act of eating) – can be argued to depend upon mental representations built on previous sensorimotor experiences with the object (Chambers, 2016; Mishra & Marmolejo-Ramos, 2010). In fact, recent years have seen a significant growth in studies examining the role of affordances, most of which have also concluded that gaze shifts towards objects in the concurrent display whose affordances are in line with the denoted action (Altmann & Mirković, 2009; Chambers, Tanenhaus, & Magnuson, 2004; Chambers, Tanenhaus, Eberhard, Filip, & Carlson, 2002; Mishra & Marmolejo-Ramos, 2010).

Affordances, however, are not only based on pairwise relationships between actions and objects (e.g., eat – cake) but also by physical characteristics of the immediate environment. In a study by Chambers and colleagues (2002), participants were given instructions to manipulate real world objects (e.g., Put the cube inside/below the can). The results revealed that the preposition inside limited referential candidates to those objects with compatible properties (i.e., containers; Chambers et al., 2002, Experiment 1). In a follow-up experiment, the size of the target object (e.g., small vs. large cube) was manipulated. Given a large cube (target object) and a display with a big and a small can, among other things, participants were found to rapidly narrow their selection to not only containers in general but to containers large enough to accommodate the cube in question. However, this behaviour was not observed when given a small cube because the object could be placed inside both containers (Chambers et al., 2002, Experiment 2). This clearly illustrates that participants draw on both linguistic (e.g., inside) and nonlinguistic (e.g., container size) information when anticipating relevant referents from linguistically-expressed actions.
In yet another study, Chambers and San Juan (2008) examined whether the verb *return* which implies a physical action but is not associated with any particular object (as you can return any object) could also influence referential interpretation. That is, lexical level associations (e.g., *eat* – cake, *inside* – containers) cannot play any significant role in linking the verb to candidate referents. In one experiment, participants were presented with a visual array that included nine labelled grid squares, four of which were occupied by objects. Participants were first asked to move an object to a specific location (e.g., *Move the chair to area two*) and then in subsequent instructions were asked to either *move* or *return* the object to another location (e.g., *Now move/return the chair to area five*). Intriguingly, the verb *return* caused immediate anticipatory eye movements to the object (e.g., chair) that was recently displaced. This demonstrates that the "affordances" of objects that drive anticipatory eye movements based on verb information are not limited to only properties that are perceptible as the verb is heard. In another line of research, Myung and colleagues (2006), further revealed that the computation of affordances matters even for nouns. Participants were more likely to fixate on the functional competitor *typewriter* than an unrelated item *bucket* when they heard the target noun *piano* (Myung, Blumstein, & Sedivy, 2006), reflecting that an action-based association (i.e., how an object is physically manipulated) could also influence the candidates that are considered as a noun is mapped to its target referent.

Previous studies clearly demonstrate that information in regards to object affordances affect early visual attention and guide on-line language processing at the same time as does the information in the speech stream, hence revealing the dynamic relationship between vision, language, and action (Heard, Masson, & Bub, 2014; Huettig, Rommers, & Meyer, 2011; Knoeferle, 2015; Mishra & Marmolejo-Ramos, 2010). In fact, a popular theory in the embodied cognition literature, the *simulation* theory postulates that an individual’s understanding of a spoken motion event involves the same sensorimotor representations activated during an individual’s physical participation in that event (Barsalou, 1999; Madden & Zwaan, 2006; Zwaan & Taylor, 2006; Taylor, Lev-Ari, & Zwaan, 2008). Some accounts have suggested that these motoric representations are responsible for the effects of action verbs on eye movement behaviour during on-line language processing (e.g., Kamide, Lindsay, Scheepers, & Kukona, 2016, Experiment 2; Heard et al., 2014). However, the degree to which the iconicity of images interacts with the information about affordances has not yet been examined in the context of a visual world study.
To date, the influence of image type has been examined mostly in relation to children’s language acquisition. As previously discussed, the extent to which children manually explore 2D images varies as a function of the iconicity of images. Infants commonly treat colour photographs as if they were real 3D objects by physically performing the relevant action (e.g., grasping a bottle) on the image itself (Pierroutsakos & DeLoache, 2003; Troseth et al., 2004). In fact, their manual interaction with the image resembles the same motion behaviour expected with the real object (e.g., grasping solid objects vs. rubbing non-solid objects). This shows that children change their behaviour to match the action afforded by the object, particularly when the object is depicted with a higher degree of iconicity (i.e., photographs). Furthermore, studies on adults have also shown that recognition for manipulable objects (which are strongly associated with action-based representations) is better when the objects are depicted as photographs than line drawings (e.g., Salmon et al., 2014). The idea that on-line language processing is influenced by both perceptual and conceptual knowledge afforded by the visual display (Altmann & Kamide, 2007; Altmann & Mirković, 2009; Chambers et al., 2002; Chambers et al., 2004; Chambers, 2016; Coco & Kellar, 2015) is the motivation behind our second experiment.

According to previous research, stimuli presented as line drawings or clipart are more likely to elicit a general abstract category concept (Bramão et al., 2011; Gelman et al., 2008), whereas a photographic image is more representative of the particular referent or exemplar (Bramão et al., 2011). Thus the relevant question is not simply how “efficiently” a referential object is identified from a noun based on the object’s status as a realistic or less-realistic image, but also the extent to which abstract category information about the object and its affordances are accessed and interpreted in the course of language processing. Affordances are influenced by the salient physical properties of an object (e.g., shape, colour, texture, orientation; Chambers, 2016), which to a great extent are the same properties that differentiate photographs and clipart images from one another. We predict that there will be greater sensorimotor activation with photograph than clipart images because they are better matched with our existing mental representations and past experiences with the real object itself. Furthermore, this type of effect would arguably be more likely to occur in a verb paradigm, on the basis of previous findings that affordances imposed by action verbs have shown to influence anticipatory eye movements in this paradigm (e.g., Altmann & Kamide, 1999; Staub et al., 2012). Therefore, Experiment 2 will examine whether
the iconicity of images would, in fact, affect anticipatory eye movement in the context of a verb paradigm, namely action verbs.

3 Experiment 2

3.1 Method

3.1.1 Participants

Similar to Experiment 1, 40 students (Age Range = 18 to 53, \( M = 22, \ SD = 7.49 \)) were recruited from the University of Toronto Mississauga to participate in the present study. Participants either received a course credit or $10 per hour as compensation. As before, participant selection was limited to individuals with native-like proficiency in English, normal hearing, and normal or corrected to normal vision.

3.1.2 Materials and procedure

Aside from the new auditory stimuli, which were recorded to incorporate the new restrictive verb (e.g., Jamie will drink the beer), all other materials and procedure were identical to the first experiment. The verbs in critical trials were changed from the non-restrictive verb move to a restrictive verb (e.g., drink, see Table 2.1) narrowing down the referent selection to the two semantically compatible objects in each display, namely a target object and the alternative (e.g., beer and wine). The extent to which the target and the alternative were related to the verb was controlled using pairwise measures of semantic association drawn from the Latent Semantic Analysis website (lsa.colorado.edu). Each verb was tailored to the objects occurring on a given critical trial. The verbs in the filler trials were also all unique; however, they varied in terms of the number of objects they were compatible with (ranging from one to all of the four displayed objects). Once again, image type (photograph, clipart) and target frequency were balanced across lists. The auditory stimuli were recorded using the same female voice and recording parameters as in Experiment 1.

3.1.3 Data analysis

The data analysis techniques were similar to Experiment 1 with two exceptions. First, the analysis region was based on the verb in the target sentence, rather than the noun. Second, verb
type (non-restrictive vs. restrictive, using data from Experiment 1 and Experiment 2 respectively) was added as a new fixed effect. The full model therefore included verb type and image type as fixed effects with items and subject as random intercepts, and item added as random slope (random slope for subject was not included as verb type was manipulated across experiments). Therefore, any analysis after the sentence onset was based on a 2 x 2 mixed design where the verb type (non-restrictive vs. restrictive) was manipulated as a between-group factor and image type (photograph vs. clipart) was treated as a within-group factor.

3.2 Results and Discussion

3.2.1 Prior to sentence onset

Recall that, in Experiment 1, there was a main effect of image type prior to sentence onset, reflecting that the number of objects that participants fixated on was greater when the stimuli were photographs, whereas the average duration of individual fixation was greater for the clipart stimuli. However, these differences were no longer observed in Experiment 2, all $ps > .05$ (see Figure 3.1). One possible explanation is that the use of restrictive verbs like peel and melt instead of the neutral verb move highlighted the function of objects, in turn drawing some attention away from the fine-grained visual properties of objects. This could potentially lead participants to engage in less active exploration of photographs, whose visual properties are more complex, eliminating the previously-observed differences in the scanning of clipart and photograph images. Although not significant, participants fixated on fewer objects overall prior to sentence onset in Experiment 2 ($M = 2.91$, $SD = .88$) than they did in Experiment 1 ($M = 3.09$, $SD = .63$), providing some suggestive evidence for this idea. However, because the actual basis of the image type effect observed in Experiment 1 is unknown (see Section 2.2.1), the explanation for the different pattern observed in the current experiment is clearly speculative.
Figure 3.1. The effect of image type prior to sentence onset for the (a) number of fixated objects, and (b) average fixation duration in Experiment 2

3.2.2 Verb region

Next we examined listeners' potential to anticipate action-compatible targets based on verb information. All the analyses in the verb region are based on the results collapsed across both experiments. This is because most likely prediction of the embodied account is an interaction between image type (photograph/clipart) and verb type (non-restrictive/restrictive), that is, that more realistic depictions should matter more for restrictive verbs like drink and peel because they evoke richer and more specific motor plans (cf. Salmon et al., 2014 in Section 1.2.2). A 1000 ms time interval was used reflecting the average duration of the verb + the (840 ms) along with a programming margin of approximately 150 ms. The interval ended 100 ms after noun onset. This interval, however, would not reflect the use of noun information because of the 200 ms saccadic margin.

Figure 3.2 plots the proportion of fixations to the two compatible targets on the visual display (e.g., wine and beer for the verb drink) as a function of image type (photograph vs. clipart) and verb type (non-restrictive vs. restrictive). Non-restrictive and restrictive verbs are denoted by Expt 1. and Expt. 2 on the figure, respectively. The measure of interest in the present experiment
was based on the proportion of fixations averaged across two objects instead of the one object in Experiment 1. Recall from Experiment 1 that the design included two alternative targets to ensure that any observed effects could not accidentally arise from a particular object being selected as the target. In Experiment 2, however, in order for two objects to be potential targets, they both had to be semantically compatible with the verb. Therefore, in any critical trial, the anticipatory effect of the verb would be reflected in looks to either of these objects (e.g., *drink* narrows selection to the semantically compatible objects *beer* and *wine*).

![Figure 3.2](image)

**Figure 3.2.** Proportion of fixations to the target and alternative target as a function of image type and verb type in Experiments 1 and 2

As expected, we observed a significant effect of verb type, whereby the proportion of anticipatory fixations to compatible referents was greater for the restrictive verb condition in Experiment 2 ($M = 4.06$, $SD = 1.78$) than the neutral verb condition in Experiment 1 ($M = 1.36$, $SD = 1.19$), $\beta = -1.36$, $SE = .17$, $t_{(63.70)} = -7.81$, $p < .001$. These results are therefore compatible with earlier studies showing verb-driven referential predictions (see Table 3.1).
Table 3.1. Main and Interaction Effects as a Function of Image Type in Experiment 2

<table>
<thead>
<tr>
<th>Effect</th>
<th>Estimate</th>
<th>SE</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before Sound: Number of Fixated Objects</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Intercept)</td>
<td>2.91</td>
<td>.14</td>
<td>21.39</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Image Type</td>
<td>-.05</td>
<td>.04</td>
<td>-1.38</td>
<td>.177</td>
</tr>
<tr>
<td>Before Sound: Average Fixation Duration (ms)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Intercept)</td>
<td>304.92</td>
<td>13.94</td>
<td>21.87</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Image Type</td>
<td>.07</td>
<td>5.76</td>
<td>.01</td>
<td>.991</td>
</tr>
<tr>
<td>Verb Region: Proportion of Fixations (^1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Intercept)</td>
<td>2.70</td>
<td>.18</td>
<td>14.93</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Image Type</td>
<td>.11</td>
<td>.11</td>
<td>.97</td>
<td>.333</td>
</tr>
<tr>
<td>Verb Type</td>
<td>-1.36</td>
<td>.17</td>
<td>-7.81</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Image x Verb Type</td>
<td>.03</td>
<td>.12</td>
<td>.26</td>
<td>.798</td>
</tr>
</tbody>
</table>

Note. \(^1\)The proportion of fixations in the verb region is based on the sum of looks to the target and semantically related object.

Although the measures in the present study were sensitive enough to detect well-established psycholinguistic effects, the main outcome of interest was still not observed. That is, there was no main effect of image type nor was there an interaction effect of image by verb type, all \(ps > .05\) (see Table 3.1). Hence, once again, the iconicity of images did not seem to have any influence on real-time language processing. Contrary to the predictions of embodied approaches to language processing, there was no benefit for photographs over clipart stimuli, which suggests that listeners treat clipart and photographic objects similarly in the course of mapping denoted actions to compatible referents in the visual environment.
Chapter 4
General Discussion

Previous studies using the visual world paradigm have revealed that listeners are sensitive to a wide range of perceptual and conceptual factors as they relate spoken language to objects in a visual environment. As already mentioned, researchers have arrived at this conclusion by employing many different kinds of visual stimuli (i.e., real objects, photographs, line drawings) that vary widely from study to study with typically no recognition or comment on these differences. However, according to the developmental and visual cognition literature, differences in terms of the level of iconicity do, in fact, matter for certain kinds of tasks. Images that are high in degree of iconicity (more realistic) have shown to benefit a number of cognitive processes such as aspects of language acquisition, concept formation, and object recognition. Furthermore, we know from past research that the ability to see images beyond their surface properties (i.e., pictorial competence) is not an innate ability but one that is learned. Thus, there are reasons to consider in greater detail what kinds of processing differences might be relevant as a function of the iconicity of images, and how these might be detected within the context of a visual world experiment.

Also, although iconicity per se has not been examined in the visual world studies, there is evidence for perceptually-driven effects in this paradigm. Salient visual characteristics of an object, particularly its shape and colour, have shown to influence the course of on-line language processing (see Huettig et al., 2011, for a review). Motivated by these past findings, the present study was designed to examine the extent to which difference in image type (photograph vs. clipart) might influence the mapping of language to referents in the immediate environment. To accomplish this, a new stimulus set was created by digitally transforming photographs of real objects to clipart analogues. Thus, all images used in the present study (photograph and clipart) represented the same real world object using a closely-matched depiction that differed in terms of their level of iconicity.

4 Does the Iconicity of an Image Matter?

From previous literature, we know that both bottom-up (the physical characteristics of the particular stimulus such as shape, colour, orientation) and top-down processes (e.g., attention,
memory, perspective) influence real-time language comprehension, particularly those relevant to goal-directed behaviour imposed by the experimental context (Chambers et al., 2004; Mishra & Marmolejo-Ramos, 2010; Salverda, Brown, & Tanenhaus, 2011). In fact, initial fixations have sometimes been argued to be driven by bottom-up processes, whereas later fixations may be controlled more by top-down processes (Mishra & Marmolejo-Ramos, 2010). Although in the current study, photographs and clipart images shared some similarity in terms of their coarse level perceptual features (e.g., orientation, scale, shape, colouring), they still differed in terms of finer grained details, thus contributing to their different level of iconicity. Photographs with more complex and detailed surface features, particularly realistic colours and shadows, were more representative of the real object. Conversely, clipart images with their clear contour and uniform colouring were considered as a more simplistic depiction of the real object. Therefore, one possibility is that photographs with their high level of iconicity would benefit the process of mapping nouns to referential objects (Experiment 1); however, this was not observed. Experiment 2 provided a stronger test by including restrictive action verbs, which are thought to drive anticipatory eye movements via sensorimotor representations of actions and affected objects. On this account, one would predict stronger verb-driven effects with photographs due to their more realistic attributes (e.g., texture and shadow information would provide stronger cues as to physical interaction with an object). Once again, the degree of the iconicity of images had no influence on the referential mapping process. Similar to previous findings, however, the present study demonstrated incrementality in the mapping of verbs and nouns to depicted objects, thus confirming the validity of our experimental paradigm.

Nonetheless, it was discovered that pictorial depiction could have an influence on the early moments of visual processing (either more efficient processing for photographs or clipart sustaining attention for longer). The observed difference between the two experiments in terms of these early effects was not expected, particularly since the task was the same until sentence onset. One (speculative) explanation offered in Section 3.2.1 was that the use of restrictive verb might have reduced the contribution of bottom-up factors during the initial visual processing because more emphasis was placed on the function of objects. Consequently, this would have directed attention away from the fine-grained visual attributes of the depicted objects. Hence, the observed difference between clipart and photograph stimuli in Experiment 1 was eliminated in Experiment 2 because participants were not as engaged in processing the visual attributes that
differentiated the two image types from one another. Additional experiments would be necessary to provide a direct test of this account.

To some extent, the present study’s results resemble the study described earlier by Wimmer and colleagues’ (2013). Recall that these authors examined whether children’s confusion about object-image relationship (e.g., inaccurate expectation that change of sticker on the doll would be transferred to its pictorial representation) would vary as a function of the iconicity of images. Children showed a similar level of confusion independent of the degree of the iconicity of images; however, they did show sensitivity to object iconicity in a follow-up simple recognition experiment. Photographs were preferentially selected as the better representation of real world object. Thus, the authors concluded that iconicity matters for perceivers’ initial understanding of object-referent relationships, but once recognition is established there is no longer an influence of image type on the given task (Wimmer et al., 2013). Accordingly, the present study’s results suggest that image type matters for the initial phase of object recognition (particularly when the materials do not encourage specific scanning strategies); however, once recognition is complete, the level of iconicity of images is no longer relevant and has no influence on the referential mapping process.

In sum, then, the results from the two experiments revealed no effect of image type (photograph vs. clipart) during the process of mapping incoming language to the referent in the visual environment. These results are, however, consistent with some other work suggesting there is no effect of iconicity. It seems that the differences in the degree of realism that distinguish clipart images from photographs (Experiments 1 and 2) or scenes (e.g., Andersson et al., 2011; Staub et al., 2012) does not affect the course of real-time language comprehension, at least given the current way in which the visual world paradigm was implemented.

4.1 Methodological Considerations and Future Directions

There are significant methodological differences across studies using the visual world paradigm. In fact, the present study tried to examine one such methodological disparity: the extent to which more or less realistic visual stimuli affect the course and nature of real-time language processing. Although we did not observe any differences as a function of image type during the course of linguistic processing, we cannot be certain that the iconicity of images would not play a
significant role in other versions of this paradigm. Alternative methodological designs, particularly those that vary in terms of the presentation and duration of visual stimuli, the complexity of auditory stimuli, as well as the experimental population could all possibly make apparent the effect of interest.

4.1.1 Stimuli

When presented with visual stimuli, the extent to which they capture our attention is influenced by both bottom-up and top-down factors, which in turn are influenced by the experimental context (Chambers et al., 2004; Hintz & Huettig, 2015; Kan & Thompson-Schill, 2004; Tatler, 2016). Although the question about the role of image type has been previously examined using richer visual contexts like scenes, there has never been a within group comparison nor has there been a controlled matching of images across studies. The present study was the first to our knowledge to examine the effect of image type within the same experimental context; however, we only examined simple arrays of objects. As already discussed, there are many variations of the visual world paradigm; thus, there is always the possibility that the effect of the iconicity of images would appear in the presence of more complex visual environment.

A study by Coco, Keller, and Malcolm (2015) showed anticipatory eye movements with photographic scenes, even when the target object was not present. For example, participants looked at the table upon hearing sandwich despite sandwich not being depicted. These results illustrate that participants use contextual scene cues during language processing (i.e., the canonical location where a sandwich would be found). Interestingly, these effects were also found in a version of the experiment using the "blank screen paradigm", in which the visual display was presented first, and is then removed and replaced by a blank screen before participants hear the accompanying recorded sentence. Fixations were still made to the appropriate X-Y coordinates on the blank screen corresponding to the location of the table surface in the previously presented photographic scene, which was similar to earlier findings using the blank screen paradigm with clipart scenes (e.g., Altmann, 2004; Altmann & Kamide, 2004). These findings confirm that individuals rely on their memory representations to some degree and that the effect of on-line language processing is not limited to the concurrent presentation of visual stimuli (Huettig et al., 2011). One possibility is that any effect of image type that relates to perceivers' internal representations would be more evident in this type of
experimental context. For example, one consequence of the "blank screen" approach may be that the influence of bottom-up processes decreases as the reliance on top-down cognitive processes (i.e., memory) increases. If so, it should be possible to detect differences in the degree to which image is encoded and represented in memory, perhaps reflecting the different scanning patterns for photographs and clipart in the first experiment. Interestingly, past research has suggested that encoding information for line drawings may be in fact be easier than for photographs (Lawson & Humphreys, 1996). Thus, in a "blank screen" context, the detailed surface characteristics of photographs may not necessarily benefit the referential mapping process because more detail demands a greater match and thus making for a more taxing process (Lam, Dijkstra, & Rueschemeyer, 2015). Therefore, it may be fruitful for future studies to examine whether using different types of visual displays that increase the demand on memory representation (e.g., blank screen paradigm) would yield an effect of image type.

Another issue relating to the stimuli is that, in the current study, we used comparatively simple auditory stimuli (e.g., Jamie will move/peel the banana), addressing only noun-triggered and verb-triggered eye movements. However, language-mediated eye movements can also be triggered by adjectives (e.g., the small/tall glass; Sedivy, Tanenhaus, Chambers, & Carlson, 1999), among other things. Thus, it cannot be confirmed that pictorial depiction would not be influential in other experimental contexts. For example, adjectives denoting physical surface features (e.g., colour, size, and shape) or evaluative features (e.g., ugly, comfortable) may draw more attention to the visual characteristics of an object, thus leading to an effect of image type. Therefore, one potential avenue for future studies is to examine whether the iconicity of images matters with a broader variety of contextual and linguistic materials.

4.1.2 Preview duration

Because the display is often presented prior to the speech stream in the visual world paradigm, participants’ visual attention at first is guided by perceptual information (Coco & Kellar, 2015; Dahan & Tanenhaus, 2004; Rommers et al., 2013). In most experiments, the preview of the display is often long enough to give participants time to explore the visual display; therefore, the process of mapping of visual input to existing mental representations is relatively complete at the time of sentence onset. Likewise, the 3 s preview duration in the current study was likely more than sufficient to process and recognize the objects on the visual display. (The number of fixated
objects was slightly below the maximum four, but we cannot rule out the possibility that the other object could have been viewed parafoveally.) This suggests that the effect of image type may matter only to the extent required for object recognition, and during tasks where demand on bottom-up processes is greater. However, once a mental representation is formed, and objects are recognized, the iconicity of images no longer plays a significant role.

The influence of preview duration in the visual world paradigm has been well illustrated in a series of experiments by Huettig and McQueen (2007) examining the effect of phonological, semantic, and perceptual competitors within a single array. For example, participants were given a sentence that contained a critical word (e.g., *Eventually she looked at a beaker that was in front of her*) while simultaneously viewing a display of four objects including a phonological competitor (e.g., beaver), a semantic competitor (e.g., fork), a perceptual competitor (e.g., bobbin), and an unrelated item (e.g., umbrella). When given enough time (1000 ms) to preview the display before the sentence onset, participants were more likely to look at the phonological competitor upon hearing the critical word, followed by perceptual, and semantic competitors. However, if participants were only given 200 ms preview window prior to the critical target, the fixation to a phonological competitor was no longer present. According to the authors, participants may not have had enough time to retrieve the name of the object during the shorter preview durations to cause the phonological competition effect. In follow-up experiments, changing the stimulus type also changed the course of the outcome. When the visual stimuli were switched from pictures to words (e.g., printed word BEAVER rather than a picture of a beaver), the competition occurred only at phonological level (and not at the semantic level) during both duration periods. However, in the absence of a phonological competitor, there was bias towards the semantic competitors for both types of display presentations, whereas the bias towards shape was only present during picture displays (Huettig & McQueen, 2007).

In a more recent study by Yee and colleagues (2011), the authors compared shape and semantic competitor effects while also manipulating the preview duration (e.g., shape condition: pizza - Frisbee; function condition: tape - glue). They discovered that shape driven competition was evident in experiments using a 1000 ms but not 2000 ms preview window, whereas this pattern of effect was reversed for the semantic competitors. They concluded that the activation of form preceded the activation of function because perceptual features are more prominent for initial
object recognition; however, once the object is recognized, the important factor is what must be done with it, involving more top-down processes. Thus it is clearly evident that language mediated eye movements are influenced by both the timing and the type of visual stimuli in the visual world paradigm (Huettig et al., 2011, Huettig & McQueen, 2007; Yee et al., 2011). One possibility is that an influence of the iconicity of images would be more prominent in the current experiments if shorter preview durations were used, particularly those less than 1000 ms, in accordance with the previous findings of perceptually-driven effects. Therefore, future studies should examine whether varying the preview duration would cause differential access to mental representations, and consequently influence the course of on-line language processing.

4.1.3 Population

Finally, the results of the present study might not be generalizable outside the population of young healthy university students from which the sample was drawn. The effect of image type might, in fact, be different for a population of younger or older adults who possess different levels of experience with visual media (e.g., newspaper, television, tablets). Effects of image type may be more evident in even younger children since as mentioned earlier, the ability to understand the dual representational status of images does not develop until the age of 24 months (DeLoache et al., 2003) and it is also around this time that children begin to show some similar real-time language processing abilities as adults (Knoeferle, 2015). It is therefore possible that the time course of referential mapping for photographs versus clipart will differ for children that are younger than the age of 24 months in comparison to older children or adults. This is especially interesting to consider in today’s society, where more and more children in certain cultures are interacting with digital media from a young age. In fact, we may even observe a different developmental trajectory of pictorial competence today than was observed by studies of young children conducted a decade ago (like DeLoache et al., 2003).

Furthermore, it may be interesting to examine the effect of image type across the adult life span. For example, one possibility is that differences related to the iconicity of objects might be more detectable in an older population due their lack of experiences with digital media coupled with their age-related cognitive decline. The direction of change could either favour photographic images as older adults are more likely to have been exposed to this type of pictorial depiction, or could favour clipart images as their simple depiction might be easier to process. In summary, our
main finding that there is no effect of iconicity of images on real-time language comprehension is not fully conclusive until more studies are conducted with other variations of the visual paradigm and further not generalizable until a more representative sample is acquired.

4.2 Conclusion

Real-time language comprehension involves the rapid integration of visual, conceptual, and auditory input (Altmann & Mirković, 2010; Chambers et al., 2002; Eberhard, Spivey-Knowlton, Sedivy, & Tanenhaus, 1995; Huettig et al., 2011), which is a continuous and bidirectional process. The mapping of language to a mental representation to a visual referent occurs alongside the process of mapping the visual input to mental representations to language (Huettig et al., 2011; Knoeferle, 2015; Mishra & Marmolejo-Ramos, 2010; Rigoli & Spivey, 2015). Therefore, the extent to which nonlinguistic (e.g., photograph vs. clipart) and linguistic (e.g., restrictive vs. non-restrictive) contexts influence the course and nature of on-line language processing depends on how efficiently these processes interact, compete, and resolve their discrepancies (Rigoli & Spivey, 2015).

The results of our two experiments are consistent with previous studies in terms of illustrating that on-line language comprehension is both incremental and anticipatory (Altmann & Kamide, 1999; Tanenhaus et al., 1995). More important, the current study found no differences related to the effect of pictorial depiction on linguistic processing. This finding is also in accordance with the handful of studies that have examined the course of on-line language processing with more ecologically valid stimuli (e.g., photograph of scenes; Andersson et al., 2011, Straub et al., 2012). These outcomes bear on certain methodological concerns that are often expressed by researchers, particularly the concern that the use of unrealistic and circumscribed visual displays might distort the natural course of language processing during this experimental paradigm (Kaiser, 2016; Tanenhaus et al., 2000). The findings from present and previous studies addressing this topic in a direct way, however, provide evidence to suggest that the course and nature of on-line language processing is more robust to methodological variations (including the abstractness of depicted objects) than previously speculated. Therefore, researchers may have the flexibility to design their visual world experiments with their preferred image type and presentation format, as long as the relevant timing issues in regards to the interval between the presentation of the stimuli and the occurrence of the language are respected.
References


Knoeferle, P. & Guerra, E. (2012). What’s non-linguistic visual context? A view from language comprehension. In R. Finkbeiner, J. Meibauer, & P. B. Schumacher (Eds.), *What is a*


Salmon, J. P., Matheson, H. E., & McMullen, P. A. (2014). Photographs of manipulable objects are named more quickly than the same objects depicted as line-drawings: Evidence that photographs engage embodiment more than line-drawings. *Frontiers in Psychology, 5*


Snow, J. C., Skiba, R. M., Coleman, T. L., & Berryhill, M. E. (2014). Real-world objects are more memorable than photographs of objects. *Frontiers in Human Neuroscience, 8*


## Appendix A

Matched Photograph and Clipart Image Pairs of Critical Targets

<table>
<thead>
<tr>
<th>Target</th>
<th>Alternative Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>Photograph</td>
<td>Clipart</td>
</tr>
<tr>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
</tr>
<tr>
<td><img src="image3.png" alt="Image" /></td>
<td><img src="image4.png" alt="Image" /></td>
</tr>
<tr>
<td><img src="image5.png" alt="Image" /></td>
<td><img src="image6.png" alt="Image" /></td>
</tr>
<tr>
<td><img src="image7.png" alt="Image" /></td>
<td><img src="image8.png" alt="Image" /></td>
</tr>
<tr>
<td><img src="image9.png" alt="Image" /></td>
<td><img src="image10.png" alt="Image" /></td>
</tr>
<tr>
<td><img src="image11.png" alt="Image" /></td>
<td><img src="image12.png" alt="Image" /></td>
</tr>
<tr>
<td><img src="image13.png" alt="Image" /></td>
<td><img src="image14.png" alt="Image" /></td>
</tr>
<tr>
<td><img src="image15.png" alt="Image" /></td>
<td><img src="image16.png" alt="Image" /></td>
</tr>
<tr>
<td><img src="image17.png" alt="Image" /></td>
<td><img src="image18.png" alt="Image" /></td>
</tr>
<tr>
<td><img src="image19.png" alt="Image" /></td>
<td><img src="image20.png" alt="Image" /></td>
</tr>
<tr>
<td><img src="image21.png" alt="Image" /></td>
<td><img src="image22.png" alt="Image" /></td>
</tr>
<tr>
<td><img src="image23.png" alt="Image" /></td>
<td><img src="image24.png" alt="Image" /></td>
</tr>
<tr>
<td><img src="image25.png" alt="Image" /></td>
<td><img src="image26.png" alt="Image" /></td>
</tr>
</tbody>
</table>
Appendix B

Norming Experiments

To ensure that all images (photographs and clipart) could be easily and accurately named and thus recognized, two online norming tasks (Name Rating and Image Naming) were completed beforehand with a separate group of participants.

**Image rating.** Two separate image rating surveys (photograph and clipart) were created using the Qualtrics online software tool (Qualtrics, Provo, UT). Fifty-six participants completed the online surveys through Amazon’s Mechanical Turk and were compensated $1.50 for their participation. Each participant was first required to complete a short language background questionnaire, and the rating task then followed. In both versions (photograph, clipart), 96 images were presented (corresponding to the four images presented in the 24 critical trials in the eye tracking experiments that follow). The order of presentation was randomized for each participant. In cases where a participant completed both versions of the online survey, only the data from their first survey was included in the analysis. The results revealed that participants’ rating agreement was quite high, as they rated the provided name to be either extremely well or very well matched with the photograph or clipart images, 94% and 91%, respectively. The images that failed to elicit the same degree of association with the provided label and also with each other (photograph and its transformed clipart image) were edited or replaced with new items.

**Image naming.** Thirty-seven participants were recruited from Amazon’s Mechanical Turk and were compensated $1.50 for completing an image naming task, also designed using Qualtrics online survey tool. Following the completion of the short language questionnaire, each participant was asked to type a name for the depicted object (photograph or clipart) in the provided textbox below the image. In the absence of a sufficient match between the name chosen by the experimenter and the name selected by majority of the participants, the name of the object in question was changed for the experimental task (e.g., *cup* was changed to *mug*). The final selection of images and their associated names were based on the results obtained from both of the norming experiments.
Follow-up image rating task. Because there were changes made to the initial set of experimental images and labels, a follow-up norming experiment (name rating) was conducted to ensure that the suitability of object names did not differ across image type (photograph vs. clipart). Fifty-one participants from Amazon’ Mechanical Turk completed the new rating task using the same 5-point scale and received $1.50 as compensation. The results revealed that, although the names for photograph images were rated slightly better than the clipart images ($M = 1.17, SD = .23$ vs. $M = 1.24, SD = .20$, where 1 denotes Extremely Well on the scale), this difference was not significant, $t_{(94)} = 1.71, p = .091$. Therefore, participants did not differentially prefer names for one image type, making it unlikely that any observed effects in the following experiments could be attributed to the fit between object names and object images.
## Appendix C

**LSA Pairwise Cosine and SUBTLEXus Frequency Measures**

<table>
<thead>
<tr>
<th>Target</th>
<th>Related</th>
<th>Verb</th>
<th>Target &amp; Related</th>
<th>Verb &amp; Target</th>
<th>Verb &amp; Related</th>
<th>Target Freq.</th>
<th>Related Freq.</th>
<th>Verb Freq.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hat</td>
<td>Sunglasses</td>
<td>Wear</td>
<td>0.10</td>
<td>0.51</td>
<td>0.21</td>
<td>64.18</td>
<td>3.94</td>
<td>109.33</td>
</tr>
<tr>
<td>Tie</td>
<td>Socks</td>
<td>Fold</td>
<td>0.35</td>
<td>0.23</td>
<td>0.13</td>
<td>44.43</td>
<td>18.27</td>
<td>8.63</td>
</tr>
<tr>
<td>Shirt</td>
<td>Scarf</td>
<td>Wash</td>
<td>0.45</td>
<td>0.34</td>
<td>0.16</td>
<td>46.37</td>
<td>4.69</td>
<td>40.73</td>
</tr>
<tr>
<td>Bag</td>
<td>Umbrella</td>
<td>Open</td>
<td>0.39</td>
<td>0.35</td>
<td>0.33</td>
<td>94.04</td>
<td>7.49</td>
<td>320.41</td>
</tr>
<tr>
<td>Shoe</td>
<td>Belt</td>
<td>Tighten</td>
<td>0.23</td>
<td>0.14</td>
<td>0.31</td>
<td>30.39</td>
<td>24.35</td>
<td>2.86</td>
</tr>
<tr>
<td>Safe</td>
<td>Chest</td>
<td>Close</td>
<td>0.10</td>
<td>0.35</td>
<td>0.28</td>
<td>143.2</td>
<td>40.98</td>
<td>219.43</td>
</tr>
<tr>
<td>Lamp</td>
<td>Flashlight</td>
<td>Turn on</td>
<td>0.68</td>
<td>0.25</td>
<td>0.32</td>
<td>12.88</td>
<td>5.92</td>
<td>306.47</td>
</tr>
<tr>
<td>Candle</td>
<td>Lantern</td>
<td>Light</td>
<td>0.45</td>
<td>0.55</td>
<td>0.52</td>
<td>8.02</td>
<td>2.02</td>
<td>165.2</td>
</tr>
<tr>
<td>Jar</td>
<td>Mug</td>
<td>Empty</td>
<td>0.14</td>
<td>0.31</td>
<td>0.28</td>
<td>8.31</td>
<td>6.84</td>
<td>47.24</td>
</tr>
<tr>
<td>Pot</td>
<td>Spoon</td>
<td>Clean</td>
<td>0.47</td>
<td>0.21</td>
<td>0.23</td>
<td>22.53</td>
<td>7.61</td>
<td>121.24</td>
</tr>
<tr>
<td>Vase</td>
<td>Pail</td>
<td>Fill</td>
<td>0.25</td>
<td>0.24</td>
<td>0.33</td>
<td>3.84</td>
<td>0.94</td>
<td>43.94</td>
</tr>
<tr>
<td>Ball</td>
<td>Dice</td>
<td>Roll</td>
<td>0.21</td>
<td>0.26</td>
<td>0.05</td>
<td>104.96</td>
<td>10.45</td>
<td>63.27</td>
</tr>
<tr>
<td>Recorder</td>
<td>Harmonica</td>
<td>Play</td>
<td>0.30</td>
<td>0.24</td>
<td>0.35</td>
<td>4.94</td>
<td>1.75</td>
<td>354.53</td>
</tr>
<tr>
<td>Present</td>
<td>Candy</td>
<td>Unwrap</td>
<td>0.10</td>
<td>0.15</td>
<td>0.09</td>
<td>89.45</td>
<td>35.78</td>
<td>0.49</td>
</tr>
<tr>
<td>Rope</td>
<td>Bow</td>
<td>Untie</td>
<td>0.35</td>
<td>0.25</td>
<td>0.19</td>
<td>22.71</td>
<td>20.27</td>
<td>5.18</td>
</tr>
<tr>
<td>Grapes</td>
<td>Strawberries</td>
<td>Rinse</td>
<td>0.37</td>
<td>0.06</td>
<td>0.09</td>
<td>3.94</td>
<td>3.92</td>
<td>2.35</td>
</tr>
<tr>
<td>Pie</td>
<td>Cookies</td>
<td>Eat</td>
<td>0.58</td>
<td>0.29</td>
<td>0.33</td>
<td>28.75</td>
<td>17.9</td>
<td>251.88</td>
</tr>
<tr>
<td>Chocolate</td>
<td>Butter</td>
<td>Melt</td>
<td>0.64</td>
<td>0.28</td>
<td>0.11</td>
<td>29.39</td>
<td>20.43</td>
<td>7.31</td>
</tr>
<tr>
<td>Beer</td>
<td>Wine</td>
<td>Drink</td>
<td>0.85</td>
<td>0.84</td>
<td>0.78</td>
<td>75.49</td>
<td>60.35</td>
<td>247.39</td>
</tr>
<tr>
<td>Onion</td>
<td>Carrot</td>
<td>Chop</td>
<td>0.31</td>
<td>0.13</td>
<td>0.24</td>
<td>4.24</td>
<td>3.24</td>
<td>13.61</td>
</tr>
<tr>
<td>Eggs</td>
<td>Noodles</td>
<td>Boil</td>
<td>0.28</td>
<td>0.12</td>
<td>0.20</td>
<td>38.63</td>
<td>6.06</td>
<td>5.94</td>
</tr>
<tr>
<td>Bread</td>
<td>Lettuce</td>
<td>Cut</td>
<td>0.59</td>
<td>0.10</td>
<td>0.20</td>
<td>28.33</td>
<td>3.39</td>
<td>229.76</td>
</tr>
<tr>
<td>Apple</td>
<td>Banana</td>
<td>Peel</td>
<td>0.32</td>
<td>0.46</td>
<td>0.30</td>
<td>23.67</td>
<td>10.73</td>
<td>5.35</td>
</tr>
<tr>
<td>Cake</td>
<td>Cheese</td>
<td>Slice</td>
<td>0.48</td>
<td>0.47</td>
<td>0.55</td>
<td>45.06</td>
<td>39.04</td>
<td>8.53</td>
</tr>
</tbody>
</table>