WORD LEARNING FROM VIDEOS: EVIDENCE FROM 2-YEAR-OLDS

by

SARAH REBEKAH ALLEN

A THESIS

Submitted in partial fulfillment of the requirements
For the degree of Master of Science
In the Department of Human Development and Family Studies
in the Graduate School of
The University of Alabama

TUSCALOOSA

2009
ABSTRACT

Young children are exposed to large amounts of screen media daily and many parents believe educational media is a source of learning for their children. However, whether or not children are able to learn information from television and video is still inconclusive. Of specific interest in the current studies was whether or not children are able to learn words from television or videos. In Study 1, 2-year olds were introduced to a novel word through a short animated video and then tested to assess their ability to comprehend and produce the name of the word in both a video setting and a live setting. Study 2 examined 2-year olds ability to honor mutual exclusivity to words learned from videos using the same experimental tasks. The results suggest children are able to add a word learned from video to their productive vocabulary, transfer their knowledge of the word learned from video into their everyday setting, and apply the word learning principle of mutual exclusivity to words learned from videos.
ACKNOWLEDGEMENTS

I would like to thank everyone who has been instrumental in helping me complete this research project. First, I would like to thank my Savior Jesus Christ for providing for me, sustaining me, and surrounding me with a supportive family and faculty to assist me in every possible way. I am thankful for my supportive husband Tom who has encouraged me to persevere to the end. I also want to thank each child care facility who opened their doors and allowed me to use their facilities. I have been extremely blessed with a wonderful department who encourages their students and aids them to succeed. I want to thank the Department of Human Development for caring for each of their students and for truly desiring each one to succeed. I would also like to thank my committee members, Dr. Maria Hernandez-Reif and Dr. Mark Klinger for their willingness to serve and their contributions to this research. Finally, I would like to thank Dr. Jason Scofield who served as the chair of my committee. He was instrumental in this research and his encouragement and expertise were invaluable. The research that is now this thesis has been the result of the work, support, and involvement of many individuals to whom I am forever grateful.
CONTENTS

ABSTRACT ...........................................................................................................................ii

ACKNOWLEDGEMENTS ...................................................................................................iii

1. INTRODUCTION ............................................................................................................1
   Overview .......................................................................................................................1

2. LITERATURE REVIEW .................................................................................................3
   Learning Words ...........................................................................................................3
   Learning from Television and Videos ........................................................................6
   Learning Words from Television and Videos ............................................................8
   Current Studies ...........................................................................................................12

3. METHODS .......................................................................................................................14
   Study 1 .........................................................................................................................14
      Participants ..............................................................................................................14
      Procedure ..............................................................................................................14
      Warm-up Phase .....................................................................................................14
      Test Phase .............................................................................................................15
   Results ..........................................................................................................................17
      Video Pointing Versus Real Pointing ....................................................................17
      Video Naming Versus Real Naming .....................................................................17
      Pointing Versus Naming ......................................................................................18
CHAPTER 1
INTRODUCTION

Overview

The American Academy of Pediatrics (AAP) (1999) has recommended that children under the age of 2 years old not be exposed to any screen media (e.g., television, computers, video games, etc.) and that preschoolers be limited to 2 hours of screen media per day. This includes examples of educational media. However, despite the AAP’s recommendation, screen media is prominent in the lives of many young children (Comstock & Scharrer, 2007). For example, infants and toddlers watch approximately 1-2 hours of television per day (Christakis, Zimmerman, DiGiuseppe, & McCarty, 2004; Rideout & Hamel, 2006; Rideout, Vandewater & Wartella, 2003) and preschoolers watch an average of 2-3 hours of television and videos per day (Huston, Wright, Marquis, & Green, 1999; Jordan & Woodward, 2001). Interestingly, most parents are not skeptical of screen media. Rather, parents tend to believe that screen media “mostly helps” their children and a large portion of parents even believe that educational television and videos are “very important” to children’s intellectual development (Rideout et al., 2003, p. 8). Given these beliefs, it is not surprising that screen media designed to target children’s development and learning are part of a booming industry that includes the Baby Einstein brand, which alone sells over $200 million worth of products each year (Bronson & Merryman, 2006; Pankratz, 2007).
Unfortunately, research on educational media often lags behind the technology and marketing of educational media. In addition, contrary to parent’s beliefs, the existing research is often inconclusive as to whether educational media are actually beneficial to children's development and learning (i.e., “mostly helps”). This is especially true for children under 3 years old (see Linebarger & Walker, 2005; Rice, Huston, Truglio, & Wright, 1990; see also Robb, Richert, & Wartella, 2009; Zimmerman, Christakis, & Meltzoff, 2007). In fact, research has been inconclusive as to whether screen media in general are even a viable source of information for children. The most obvious consequence of inconclusive results is that parents and children may be committing significant resources (e.g., money, time, attention, etc.) to something with an unknown educational benefit instead of committing those same resources to something with a known educational benefit. In the end, more research is needed to clearly and accurately determine whether screen media, and in particular educational media, can benefit children’s development and learning. With this in mind, the current studies aimed to look more closely at one common example of learning from an educational media source - that of learning words from videos.
Learning words is one thing that young children tend to do very well. Infants usually comprehend and produce their first words before their first birthdays (Bates, Bretherton, Beeghly-Smith, & McNew, 1982; Benedict, 1979). And although word learning is initially slow (1-2 words per week), by their second birthday, toddlers are adding words to their vocabulary at a much faster rate (as many as 10-20 words per week). This phenomenon is often referred to as the vocabulary spurt (Bloom, 1973; Ganger & Brent, 2004). The result for many preschoolers is a large, rapidly expanding vocabulary (Fenson, Dale, Reznick, Bates, & Thal, 1994).

Young children’s ability to learn individual words rapidly, often after only limited exposure to those words, is known as fast-mapping (Carey & Bartlett, 1978, Heibeck & Markman, 1987). Carey and Bartlett first demonstrated fast-mapping during a task in which a child sat across the room from two objects, one that was a known color (i.e., red) and one that was an unknown color (i.e., olive-green). The experimenter then asked the child to the retrieve the “chromium” object. Carey and Bartlett found that children tended to retrieve the olive-green object and that children maintained this tendency over a delay. Carey and Bartlett (1978) called this tendency fast-mapping because children had quickly formed a hypothesis about the meaning of the new word (i.e., “chromium” meant olive-green) and retained that hypothesis over time (see Heibeck & Markman, 1987 for fast-mapping of non-color words).
Fast-mapping could be considered surprising, especially in view of the potential difficulties inherent in learning a word (Quine, 1960). The philosopher Quine used an analogy to illustrate the challenges in learning a word. In Quine's analogy, an explorer in a foreign country hears a native exclaim “gavagai” as a rabbit crosses their path. Quine suggested that, for the explorer, _gavagai_ had an endless number of possible meanings (e.g., rabbit, run, jump, food, etc) and consequently that it could be difficult for the explorer to correctly learn the meaning of _gavagai_. The implication of this analogy is that learning a word, and ultimately building a vocabulary, could be a slow and difficult process for children. However, because learning words is not a slow and difficult process, many present-day researchers believe that children must have ways of narrowing the set of possible meanings a word might have (Behrend, 1990; Markman, 1990).

One way in which children could narrow the set of meanings for a word is by relying on assumptions about how words are used (Behrend, 1990; Markman, 1990; Markman & Wachtel, 1988; Merriman & Bowman, 1989). Two such word learning assumptions are the principles of mutual exclusivity (the assumption that different objects have different names) and taxonomy (the assumption that similar objects have the same name). The principle of mutual exclusivity explains tasks in which children are presented with two objects, one familiar (e.g. a cup) and one unfamiliar (e.g. a gyroscope), and are asked to select the object that matches a word (e.g. “Can you point to the gyroscope”)? On this task, even children as young as 18 months old tend select the unfamiliar object (Markman & Wachtel, 1988; Merriman & Bowman, 1989; Scofield & Behrend, 2007; Markman, Wasow, & Hansen, 2003). Merriman and Bowman (1989) have called this tendency the disambiguation effect because it helps to reduce the ambiguity of a word’s meaning.
The principle of taxonomy explains tasks in which children are introduced to an object (e.g., “This is a clem”) and asked to select other object(s) that match a word (e.g., “Can you help me find the clem”)? On this task children (e.g., 2- to 4-year-olds) tend to select the exemplars of the original object over unrelated distracters (Behrend, Scofield, & Kleinknecht, 2001; Golinkoff, Mervis, & Hirsh-Pasek, 1994; Markman & Hutchinson, 1984). Overall, lexical assumptions like mutual exclusivity and taxomony are thought to help children narrow the set of possible meanings for a new word.

Another way in which children could narrow the set of meanings for a new word is by relying on assumptions about how speakers communicate (Baldwin, 1995; Bloom, 2000; Tomasello, 1995). Children learn many words while interacting directly with a speaker. During these interactions, speakers regularly provide important social or pragmatic cues that assist children in correctly identifying the referent of the word (i.e., the speaker’s referential intent) (Akhtar, Jipson, & Callanan, 2001; Akhtar, Carpenter, & Tomasello, 1996) and children assume that these cues are meaningful for communication. Examples of these cues include joint attention, eye gaze, pointing and other gestures, emotions, and intonations (Baldwin et al., 1996; Baldwin & Moses, 2001; Houston-Price, Plunkett, & Duffy, 2006). Under most conditions, these social cues are unambiguous and spotlight a particular referent and children tend to favor the spotlighted referent as the likely referent of the word. For example, children identify the direction of a speaker's gaze and then assume that the referent is likely to be in line with the gaze. Overall, social cues like gaze, pointing, and joint attention help narrow the set of possible meanings for a new word. Indeed, social cues are so prevalent and so helpful that some researchers believe that they are a critical part of the word learning process (Baldwin, 1991; Baldwin et al., 1996; though see Akhtar, 2005; Scofield, 2005; Scofield & Behrend, 2007).
There is a general agreement that both lexical and social or pragmatic assumptions play a role in word learning (Diesendruck & Markson, 2001; Hollich, et al., 2000). However, there is some disagreement as to the relative importance and contribution of each (Diesendruck & Markson, 2001; Jaswal & Hansen, 2006; Scofield & Behrend, 2007) - especially when it comes to which set of assumptions better or more often explains word learning successes. But, because the two sets of assumptions often converge on a single referent, differentiating between these as explanations for word learning and resolving this disagreement has been difficult. Interestingly however, screen media like television and video may provide a unique framework for differentiating between these two explanations, in part because screen media are not social in nature and, as such, are less likely to support social or pragmatic assumptions.

**Learning from Television and Videos**

Screen media are prominent in the lives of young children. However, research on the effectiveness of screen media like television and videos as good sources of information is often mixed (Barr & Hayne, 1999). For example, in examining infants learning of emotional reactions, Mumme and Fernald (2003) had 12-month-olds watch the experimenter react emotionally to a televised object while a real example of the object was visible. Mumme and Fernald found that the emotional reactions toward the televised object influenced the 12-month-old’s later behavior toward the real object. In a similar study, Strouse, Verdine, Milam, and Troseth (2008) also had 12-month-olds watch an experimenter react emotionally to a televised object while a real example of the object was not visible. Strouse et al. found that 12-month-olds had difficulty encoding emotional information about the televised object and then later applying that information to a real object. In these two examples, similar studies (i.e., studies that examined similar ages within the same domain) yielded different results.
The same is true in other domains as well. For example, in examining toddler's problem-solving, Schmidt, Crawley-Davis, and Anderson (2007) had 2-year-olds try to retrieve an object that they had previously seen hidden on a video. To account for children’s difficulty in transferring information from a two-dimensional source (i.e. video screen) to a three-dimensional location (see also Carver, Meltzoff, & Dawson, 2006; Johnson, Younger & Furrer, 2005; Zack et al., 2008), Schmidt et al. used a felt board to create a two-dimensional replica of the original video screen (see also Troseth & DeLoache, 1998). However, even with 2-dimensional replica available, children still performed poorly on the retrieval task. In examining toddler's imitation, Hayne, Herbert, and Simcock (2003) had 2-year-olds try to imitate a 3-step action sequence (e.g. Step 1 = put bell through hole into red ball; Step 2 = put plug into hole in red ball; Step 3 = shake stick to make noise). Hayne et al. (2003) found that even after a 24-hour delay, 2-year-olds were successfully able to imitate the complex action sequence seen on the video at a greater rate than children in a control condition. In these two examples, similar studies (i.e., studies that examined similar ages in related domains) again yielded different results.

While mixed results like these are not uncommon, the overall trend in research seems to show that young children are capable of learning some information from television and video sources (see Anderson & Pempek, 2005 for discussion). Yet, as is clear from the examples just reviewed, young children frequently encounter difficulties when learning from television and videos and any ability to resolve these difficulties seems to be fragile and to vary across (or even within) age, task, and domain. One clear deficit that is particularly challenging for young children comes in transferring information from a video to real life. This well-documented deficit has been labeled the video deficit effect (see Anderson & Pempek, 2005; Barr, Muentener, & Garcia, 2007; Kirkorian, Wartella, & Anderson, 2008) and is routinely found in children younger
than 3 years old whereas children older than 3 years old usually have fewer problems transferring information from a video to real life.

There are various explanations for the video deficit effect. One explanation suggests that television and videos are “perceptually impoverished” when compared to real objects, therefore younger children simply do not process the information on a video and information in real life similarly (see Anderson & Pempek, 2005 for review). Consequently, children are put at a disadvantage when trying to process one type of information that is unhelpful for identifying another type of information. Perhaps the most obvious and noteworthy perceptual difference between a video object and a real object are the number of dimensions that represent each. Past research has shown that the two-dimensional nature of video often limits young children’s ability to make three-dimensional conclusions (Schmidt et al., 2007; Zack et al., 2008).

Another explanation for the video deficit effect suggests that young children do not easily comprehend the dual nature of symbols and regularly confuse symbols and their referents (DeLoache, 1991; Troseth, Pierroutsakos, & DeLoache, 2004). Consequently, any symbolic representation, including a video, is susceptible to being confused with objects in the real world. Not surprisingly, the ability to transfer from two dimensions to three dimensions and to comprehend the dual nature of symbols tends to exempt older children from the video deficit effect (Anderson & Pempek, 2005).

**Learning Words from Television and Videos**

Interestingly, one area in which the video deficit effect might be less apparent is in the area of children’s word learning (e.g., Krcmar, Grela, & Lin, 2007). Using an experimental design, Krcmar et al. (2007) recently examined children’s word learning across four separate conditions: (a) a joint reference condition where the experimenter labeled a live target object as it
was the focus of the child’s attention, (b) a discrepant labeling condition where the experimenter labeled a live target object while the child was distracted (see Baldwin, 1991), (c) an adult-on-television condition where a child viewed a video of an experimenter labeling a target object, and (d) a *Teletubbies* condition where a child viewed a television clip of a target object being labeled. After the target object was labeled in each condition, toddlers (i.e., ranging from 15 to 24 months old) were asked to identify the correct target object from a set of five real objects, the target and four distracters. Although the toddlers tended to learn words best in the joint reference condition (i.e., correctly chose the real target object), word learning did occur in other conditions as well. The results of the last two conditions (i.e., adult-on-television and *Teletubbies*) are particularly interesting given that the video deficit effect should have negatively affected the 15- to 24-month-olds but seemingly did not.

Strouse, Saylor, Troseth, and O’Doherty (2008) also looked at children’s ability to transfer words learned from video into a real setting by looking at 30-month-olds’ ability to use social cues presented on a video to learn words from videos. Strouse et al. used a researcher (either in a live condition or video condition) to label an object in an opaque bucket with a novel word (e.g. “modi”) and to comment on another object that children could see in a transparent container. Another researcher then showed the children two real objects and asked them to pick the “modi.” The same procedure was used with two more groups of children; however, these two groups began the study with interactive games (either through live interaction or video interaction) to highlight the social cues present in the video. In both groups, Strouse et al. found that children learned the word more often in the live conditions than in the video condition. These results suggest that children may have some difficulty transferring words learned from video to a real setting.
As evidenced by the previous examples, and like other examples of learning from video, the results as to whether young children are able to learn words from video in general are mixed (e.g., Rice, Huston, Truglio, & Wright, 1990; Robb, Richert, & Wartella, 2009; Strouse, Saylor, Troseth, & O’Doherty, 2008; Zimmerman et al., 2007). Furthermore, studies that examine young children’s ability to learn words from video tend to introduce the words and test for word learning in the same medium (e.g., a novel word introduced on video and children’s learning of that word tested on video) (Rice & Woodsmall, 1988; Scofield, Williams, & Behrend, 2007; Scofield & Williams, 2009). Therefore, these studies do not directly test the video deficit effect in word learning and leave open the question of whether there is reason to believe that word learning might be less affected by the video deficit effect.

Scofield et al. (2007) recently found that children can successfully learn words from computer videos. Scofield et al. presented 2-year-olds with short animated videos in which a target object appeared on screen and was labeled by an audio track rather than by a speaker (e.g. “This is a koba”). The target then disappeared and reappeared along with three distracter objects and children were asked to select the object that best corresponded to the original word (e.g. “Can you point to the koba”)? Scofield et al. found that children overwhelmingly selected the target and suggested that word learning could be successful in the absence of a speaker or in the absence of the social cues often provided by a speaker. In addition, Scofield et al.’s results suggest that, when introduction and testing both occur on video, 2-year-olds can successfully learn words from videos. However, Scofield et al.’s results do not inform research on the video deficit effect.

In a similarly designed follow-up study, Scofield and Williams (2009) found that, despite the absence of a speaker and speaker-provided cues, words learned from videos did not simply
reduce to associations between the word and an object. Instead, words learned from videos were eligible for some of the same assumptions as words learned from speakers. Scofield and Williams presented 2-year-olds with three separate conditions: a baseline condition, a disambiguation condition, and an extension condition. The baseline condition was identical to the Scofield et al. (2007) study. In the disambiguation condition children were first presented with a target object (e.g., ‘This is a zav. A zav. This is a zav’) and were then asked to select which of two objects (i.e., the target and a distracter) best matched a new word (e.g., ‘Can you help me find the lug? Which one is the lug’)? In the extension condition, children were first presented with a target (e.g., ‘This is a clem. A clem. This is a clem.’) and were then asked to select which of two objects (i.e., an exemplar of the target and a distracter) best matched a new word (e.g., ‘Can you help me find the clem? Which one is the clem’)? Scofield and Williams found that object selection in the disambiguation condition did not differ from chance whereas object selection in the extension condition did. Scofield and Williams suggested that 2-year-olds honored the taxonomy assumption, but not necessarily the mutual exclusivity assumption, when learning words from videos.

That 2-year-olds failed to disambiguate words learned from videos was somewhat surprising given that word learning assumptions like mutual exclusivity are generally believed to be an immediate outcome of learning a word. However, Scofield and Williams (2009) offered two potential reasons for failed disambiguation: a) the few number of trials (i.e., 2) or b) the absence of social or pragmatic cues often thought to be important for resolving the disambiguation task.
Current Studies

The current studies were designed to address three questions that previous research have left unanswered or unclear about learning words from television or video. Answers to these questions will help determine the educational value of screen media sources like television and video for building young children’s vocabulary.

First, Study 1 and Study 2 ask whether the video deficit effect is apparent when children under the age of 3 years old learn words from videos. Thus far, only the results of the Krcmar et al. (2007) study suggest that toddlers may be immune to the video deficit effect when learning words from videos, but the video deficit effect was not a central focus of the study. Other studies either suggest that the video deficit effect is apparent in toddlers (e.g., Robb et al., 2009) or have used designs that do not require toddlers to transfer knowledge from video to real life (e.g., Rice & Woodsmall, 1988; Scofield et al., 2007).

Second, Study 1 and Study 2 ask whether evidence of successful word learning is only available using comprehension tasks (e.g., pointing or looking time) or whether evidence might also be available using production tasks (e.g., naming). Thus far, young children’s word learning from screen media like television and video has most often been assessed through later vocabulary growth (Linebarger & Walker, 2005; Rice et al., 1990) or through comprehension tasks (Rice & Woodmsall, 1988; Scofield & Williams, 2009; Scofield et al., 2007). However, an important aim of educational videos that target language development (and certainly the hope of the parents who purchase them) is to build children's vocabularies. In addition, past research shows that children are able to add words learned from a speaker to their productive vocabularies (e.g., Akhtar et al., 1996).
Third, Study 2 asks whether children's inability to disambiguate in the Scofield and Williams (2009) study might be explained by the few number of trials used. When learning words from speakers, children often make immediate assumptions about how those words can be used (Behrend, 1990; Markman, 1990). However, these assumptions do not tend to depend on the source of the word or on a history of experience with that source. What is puzzling about the Scofield and Williams (2009) results is that children seem to have no problem learning the word from a video (see also Scofield et al., 2007) and no problem extending that word to previously unseen exemplars. This leaves open the question of why learning a word might lead successfully to one assumption but not to another. Although trial number is not a theoretically interesting reason for failed disambiguation, it could be procedurally interesting and, if it is determined not to explain the disambiguation pattern found in Scofield and Williams, then the door is open for future studies to look more closely at possible theoretical explanations (e.g., Diesendruck & Markson, 2001).
CHAPTER 3
METHODS

Study 1

Participants

Thirty-two 24- to 36-month-old children ($M = 30$ months) were recruited from Tuscaloosa, Alabama daycares with parental consent and children’s assent. This total does not include seven children who did not pass the warm-up trials. Reasons for not passing included no participation, inattentiveness, and/or incorrect responses to all questions during the warm-up phase.

Procedure

Each participating child completed two phases: 1) a warm-up phase, consisting of two trials, and 2) a test phase, consisting of four trials. Each trial asked children to perform four specific tasks 1) pointing to a video image, 2) naming a video image, 3) pointing to a real image, and 4) naming a real image. During the session, the child and the experimenter sat at a small table in/near the child’s preschool classroom and completed a 10-minute session where the child observed a series of short animated videos created with Flash Macromedia Software displayed on a laptop computer.

Warm-up Phase

The warm up phase was designed to acclimate children to the experimenter and to the experimental procedure and used familiar images, familiar objects, and familiar words. During
both warm-up trials, children were presented with short animated videos which displayed a single image on the screen for approximately 10 seconds while an audio track simultaneously labeled the image with a word (i.e. “This is an apple”).

Children then saw the target image disappear and reappear with a distracter image (i.e. scissors) for approximately 10 seconds. Children were then asked by the audio track to point to the image that best corresponded to the target word (e.g. “Can you point to the apple”)? Following the selection, the distracter image disappeared (leaving the target image alone on the screen) and children were asked, again by the audio track, to say the word for the target image (e.g., “What is this called”)?

The experimenter then placed two objects on the table, one real example of the target image and one real example of the distracter image (i.e. a real apple and real scissors), and asked the child to point to the object that best corresponded to the target word (i.e. “Can you point to the apple”)? The experimenter then removed the distracter object (leaving the target object alone on the table) and children were asked to say the word for the target object (e.g., “What is this called”)?

Children who passed both warm-up trials ($n = 32$) continued to the test phase. Children who did not pass both warm-up trials ($n = 7$) were thanked and escorted back to their preschool classroom.

**Test Phase**

The test phase, consisting of four trials, was designed to measure children’s ability to learn (through pointing and naming) unfamiliar words and objects introduced from a video. The novel words were invented words that were unknown to children. The objects and distracters were also unknown to children and were real examples of the images that were observed on the
animated videos. Like both warm-up trials, each test trial began with a short animated video (see Figure 1) displaying a single image on the screen for approximately 10 seconds while an audio track simultaneously labeled the image with an unfamiliar target word three times (i.e. “This is a fep! It’s a fep! This is a fep”):)

The target image disappeared and then reappeared along with three distracter images, aligned in two rows of two. The images remained on the screen for approximately 10 seconds while an audio track asked the child to point to the image that best corresponded to the target word (i.e. “Can you point to the fep”? All images disappeared and the target image reappeared for approximately 10 seconds while an audio track requested a name for it (e.g. “What is this called”)?

The experimenter then placed the real objects, a real example of the target image (i.e. the target object) and real examples of the distracter images, on the table aligned in two rows of two. The objects remained in view for approximately 10 seconds while the experimenter asked the child to point to the object that best corresponded to the target word (e.g. “Can you point to the fep”? The experimenter then removed the distracter objects (leaving the target object alone on the table) and requested a word for the target object (e.g., “What is this called”? After the child made a selection, the target object was removed and the trial was completed.

Presentation orders were counterbalanced such that ½ of participants completed the video testing first and ½ completed the real testing first. In addition, ½ of participants were asked to point first and ½ were asked to name first. Also, exposure times were equal across both the video and real test phases and object position during selection of the images/objects between the video testing and the real testing were counterbalanced.
Results

In Study 1, children were presented with 4 separate videos in which a novel target image was labeled with a novel word. The aim of Study 1 was to determine: 1) whether children could identify, either through pointing or naming, the target images and target objects and 2) whether children performed differently when pointing and naming.

Video Pointing Versus Real Pointing

To determine whether children could successfully point to the target images and the target objects, the rate of pointing found in Study 1 was compared to the rate of pointing that was predicted by chance. Because there were 4 selection choices for each trial, chance was calculated as 1 out of 4 (i.e., 25%). A one sample t-test revealed that children successfully pointed to the target images ($M = 2.38$ or 60%) at a rate greater than chance, $t(31) = 6.89, p < .01$. This rate was similar to the rates found in previous studies (Scofield et al., 2007). A one sample t-test also revealed that children successfully pointed to the target objects ($M = 2.06$ or 52%) at a rate greater than chance, $t(31) = 4.40, p < .01$. Finally, a paired samples t-test revealed that the rate of pointing to the target images and the rate of pointing to the target objects did not differ, $p > .05$.

Video Naming Versus Real Naming

Children’s ability to successfully name the target images and the target objects in Study 1 could not be compared to chance since there were an unlimited number of potential names a child could produce. However, out of 4 trials children correctly produced the name of 43% of the target images ($M = 1.72$) and 46% of the target objects ($M = 1.84$). A paired samples t-test revealed that the rate of naming the target images and the rate of naming the target objects did not differ, $p > .05$. 
**Pointing Versus Naming**

To determine whether there were differences between children’s pointing and naming, the rate of successful pointing was compared to the rate of successful naming. A paired samples t-test revealed that children successfully pointed ($M = 4.44$) at a greater rate than children successfully named ($M = 3.56$), $t(31) = 2.22$, $p < .05$. However, it could be argued that the inclusion of chance influenced the scores. Therefore, the rate of successful pointing (recalculated without chance) was compared to the rate of successful naming. Another paired samples t-test revealed that there was no difference between the rate of successful pointing ($M = 3.38$) and the rate of successful naming ($M = 3.56$), $t(31) = -.424$, $p > .05$.

**Discussion**

In Study 1, children watched a video in which a novel target image was labeled with a novel word. Children were then shown a set of novel images or a set of novel objects and were asked to identify which of those images/objects best corresponded to the presented word. Of interest was whether children would successfully point to and produce the name of the unfamiliar images/objects.

The results of Study 1 indicated that children successfully pointed to both the target images and the target objects in response to the presented novel words. Children pointed successfully on 60% of the video trials and on 52% of the real trials; both percentages were significantly higher than chance (i.e. 25%). These results suggest that children can learn a novel word, as indicated by pointing, from a video and that children can transfer this learning from a video to a real setting. The results of Study 1 also indicated that children named both the target images and the target objects on 43% of the video trials and on 46% of the real object trials; both percentages show that children are producing nearly $\frac{1}{2}$ of the words presented on the video.
These results suggest that children can learn a novel word, as indicated by naming, from a video and that children can transfer this learning from a video to a real setting. Finally, the initial results of Study 1 indicated that, overall, the rate of successful pointing was higher than the rate of successful naming. Later results however suggested that pointing and naming may not differ suggesting that children who learn a word from video, point to and name the word at the same rates.

Overall, these results tend to align with the results of previous studies. For example, Scofield et al. (2007) found that young children (2-years old) successfully learned a word from a video at rates greater than predicted by chance as indicated by pointing (see also, Krcmar et al., 2007). In addition, previous studies have shown that older children (3-and 5-year olds) could successfully produce names learned on a video (Linebarger & Walker, 2005; Rice & Woodsmall, 1988; Akhtar, Carpenter, & Tomasello, 1996). Furthermore, it is not surprising that children pointed at a greater rate than naming since pointing is an easier task that involves a simpler recognition-based memory process rather than a recall-based memory process. Overall, Study 1 indicates children can learn words from a video and then later identify the words in their natural environment.

Still unknown, however, is whether children make the same basic assumptions about a word learned from videos as they do about a word learned from a speaker (i.e., the typical word learning setting). Previous studies have shown that when children learn words from speakers, they assume (among other things) that the words are mutually exclusive (Diesendruck & Markson, 2001; Markman & Wachtel, 1988; Merriman & Bowman, 1989; Scofield & Behrend, 2007). Scofield and Williams (2009) have recently addressed this question and their results indicated that children do not honor mutual exclusivity on words learned from videos. However,
Scofield and Williams suggested that the small sample size and number of trials might explain children’s performance. Study 2 was designed to more closely examine this possible explanation.

**Study 2**

**Participants**

Thirty-two 24- to 36-month-old children ($M = 33$ months) were recruited from Tuscaloosa, Alabama daycares with parental consent and children’s assent. This total does not include 8 children who did not pass the warm-up trials. Reasons for not passing included no participation, inattentiveness, and/or incorrect responses to all questions during the warm-up phase.

**Procedure**

The procedure used in Study 2 was very similar to the procedure used in Study 1. Children again completed two phases 1) a warm-up phase, consisting of two trials, and 2) a test phase, consisting of four trials. Each trial asked children to perform four specific tasks 1) pointing to a video image, 2) naming a video image, 3) pointing to a real image, and 4) naming a real image.

**Warm-up Phase**

The warm up phase was designed to acclimate children to the experimenter and to the experimental procedure and used familiar images, familiar objects, and familiar words. During both warm-up trials, children were presented with short animated videos which displayed a single image on the screen for approximately 10 seconds while an audio track simultaneously labeled the image with a word (i.e. “This is an apple”).

Children then saw the original image disappear and reappear with a distracter image (i.e. scissors) for approximately 10 seconds. An audio track then asked the children to point to the
image that best corresponded to a new word (e.g. “Can you point to the scissors”)? Following the selection, the original image disappeared (leaving the distracter image alone on the screen) and an audio track asked children to say the word for the distracter image (e.g., “What is this called”)?

The experimenter then placed two objects on the table, one real example of the original image and one real example of the distracter image (i.e. a real apple and real scissors) and asked the child to point to the object that best corresponded to the new word (e.g. “Can you point to the scissors”)? The experimenter then removed the original object (leaving the distracter object alone on the table) and children were asked to say the word for the distracter object (e.g., “What is this called”)?

Children who passed both warm-up trials ($n = 32$) continued to the test phase. Children who did not pass both warm-up trials ($n = 8$) were thanked and escorted back to their preschool classroom.

**Test Phase**

The test phase, consisting of four trials, was designed to measure children’s ability to apply mutual exclusivity to words introduced from a video similarly to the way they use words learned from a direct interaction. Like both warm-up trials, each test trial began with a short animated video (see Figure 2) displaying a single image on the screen for approximately 10 seconds while an audio track simultaneously labeled the image with an unfamiliar word three times (i.e. “This is a vox! It’s a vox! This is a vox”!

The original image disappeared and then reappeared along with one distracter image, aligned side by side. The images remained on the screen for approximately 10 seconds while an audio track asked the child to point to the image that best corresponded to a new word (i.e. “Can
The experimenter then placed the real objects, a real example of the original image and a real example of the distracter image, on the table aligned side by side. The objects remained in view for approximately 10 seconds while the experimenter asked the child to point to the object that best corresponded to the new word (i.e. “Can you point to the hap”)? The experimenter then removed the original object (leaving the distracter object alone on the table) and requested a word for the distracter object (i.e., “What is this called”)? After the child made a selection, the distracter object was removed and the trial was complete.

Presentation orders were counterbalanced such that ½ of participants completed the video pointing and naming first and ½ completed the real pointing and naming first. Also, exposure times were equal across both the video and real phases and object position of the images/objects between the video phase and real phase were counterbalanced.

Results

The procedure used in Study 2 was similar to the procedure used in Study 1. The aim of Study 2, however, was to determine: 1) whether children could identify, again through pointing or naming, new novel images and new novel objects and 2) whether children performed differently when pointing and naming.

Video Pointing Versus Real Pointing

To determine whether children could successfully point to the new target images and the new target objects, the rate of pointing found in Study 2 was compared to the rate of pointing that was predicted by chance. Because there were 2 selection choices for each trial, chance was
calculated as 1 out of 2 (i.e., 50%). A one sample t-test revealed that children successfully pointed to the new target images \((M = 3.03 \text{ or } 76\%)\) at a rate greater than chance, \(t(31) = 6.25, p < .01\). A one sample t-test also revealed that children successfully pointed to the new target objects \((M = 2.81 \text{ or } 70\%)\) at a rate greater than chance, \(t(31) = 5.35, p < .01\). Finally, a paired samples t-test revealed that the rate of pointing to the new target images and the rate of pointing to the new target objects did not differ, \(p > .05\).

**Video Naming Versus Real Naming**

As in Study 1, because there were an unlimited number of potential names for each image or object, the results for naming could not be compared to chance. However, out of 4 trials children correctly produced the name of 49% of the new target images \((M = 1.94 \text{ or } 49\%)\) and 56% of the new target objects \((M = 2.22 \text{ or } 56\%)\). A paired samples t-test revealed that the rate of naming the new target images and the rate of naming the new target objects did not differ, \(p > .05\).

**Pointing Versus Naming**

To determine whether there were differences between children’s pointing and naming, the rate of successful pointing was compared to the rate of successful naming. A paired samples t-test revealed that children successfully pointed \((M = 5.84)\) at a greater rate than children successfully named \((M = 4.16), t(31) = 3.79, p < .01\). However, it could be argued that the inclusion of chance influenced the scores. Therefore, the rate of successful pointing (recalculated without chance) was compared to the rate of successful naming. Another paired samples t-test revealed that there was no difference between the rate of successful pointing \((M = 3.94)\) and the rate of successful naming \((M = 4.16), t(31) = -.486, p > .05\).
Discussion

In Study 2, children again watched a video in which a novel image was named with a novel word (i.e. “This is a vox”). Children were then shown a pair of images, the original image and a new novel image, or a pair of objects, the original object and a new novel object, and were asked to identify the image or object that best corresponded to a new novel word (“Can you point to the hap”)? Of interest was whether children would successfully point to and name the new target image or object.

The results of Study 2 indicated that children successfully disambiguated. That is, children successfully pointed to both the new target images and the new target objects in response to the new novel word. Children pointed successfully on 76% of novel image trials and on 70% of novel object trials; both percentages were significantly higher than chance. These results suggest that children can successfully apply the principle of mutual exclusivity to a newly learned word from a video, as indicated by pointing, and that children can transfer this learning from a video to real life.

The results of Study 2 also indicated that children named both the new target images and the new target objects on 49% of novel image trials and on 56% of novel object trials; both percentages show that children are producing nearly ½ of the words presented on the video. These results again suggest that children can apply the principle of mutual exclusivity to a newly learned word from a video, as indicated by naming, and that children can transfer this learning from a video to real life. Finally, and like Study 1, the initial results of Study 2 indicated that, overall, the rate of successful pointing was higher than the rate of successful naming. Later results however suggested that pointing and naming may not differ suggesting that children who learn a disambiguated word from video, point to and name the word at the same rates.
Overall, these results differ from the results of Scofield and Williams (2009). Recall that Scofield and Williams found that children honored taxonomy, but not mutual exclusivity, when learning a word from video. Although the studies were designed similarly, Study 2 and the Scofield and Williams study differed in the sample size and in the number of trials and the number of tasks (e.g., Study 2 included both pointing and naming) children completed. Overall, Study 2 suggests that children honor the principle of mutual exclusivity on words learned from a video and successfully transfer this knowledge from the video to real life.
CHAPTER 4
OVERALL DISCUSSION

Previous studies had shown that children could successfully learn words from screen media sources such as television, videos, and computers (Krcmar et al., 2007; Rice & Woodsmall, 1988; Rice et al., 1990; Scofield et al., 2007) and even that children could honor some assumptions about how newly learned words are used (Scofield and Williams, 2009). Unclear from these studies however was: a) whether children could reliably transfer their knowledge of a word learned from video to real life (i.e., avoid the video deficit effect), b) whether children could add a word learned from video to their productive vocabulary, and c) whether children could reliably disambiguate a word learned from video. The current studies aimed to address these issues.

In Study 1, 2-year-olds watched animated videos in which a novel object image was labeled with a novel word. Children were then asked to show evidence of word learning by pointing to and naming both the original novel image and its real life counterpart. In Study 2, 2-year-olds watched similar videos but were asked to show evidence of disambiguation by pointing to and naming a new novel image and its real life counterpart. The results of Study 1 indicated that children successfully learned a new word from video, as evidenced by both pointing and
The results of Study 2 indicated that children successfully disambiguated a new word learned from a video, again as evidenced by both pointing and naming. Finally, the results of both Study 1 and Study 2 indicated that children successfully transferred their knowledge of the newly learned or newly disambiguated word from a video image to the real life counterpart (i.e., children avoided the video deficit effect). Overall, children’s ability to learn, disambiguate, produce, and transfer knowledge of a new word from video suggests that screen media like television and video can be effective sources for children learning words.

As previously stated, young children often learn words while interacting directly with a speaker and generally rely on the social cues that are provided during those interactions (Baldwin & Moses, 2001; Houston-Price et al., 2006; Tomasello, 1995). The most common social cues (e.g., joint attention, pointing, and gaze) are thought to help support word learning by focusing the child’s attention on the object or event in the world that the speaker intends to label (Baldwin et al., 1996; Baldwin & Moses, 2001; Tomasello, 1995). Consequently, children can use these cues to narrow the set of possible meanings a new word might have thereby simplifying the task of learning the meaning of that particular word.

An important difference between the typical examples of word learning and the examples found in some trials of the current studies was the absence of a speaker and therefore a

---

1 Although performance on the naming task could not be compared to chance, the current studies did find that children were able to produce roughly 50% of the words introduced on the video suggesting that after only limited exposure, children had the ability to add some of the words to their productive vocabulary. It is certainly noteworthy that children in the current studies were better at correctly pointing (i.e., comprehension) than at correctly naming (i.e., production), but it is not necessarily surprising (see Akhtar et al., 1996). Correctly naming the objects required children to rely on recall memory which is more difficult than relying on recognition memory because it involves a more elaborate retrieval process (Bellack, 1963; Calvert, Huston, Watkins, & Wright, 1982; Naus, Omstein, & Kreshtool, 1977). This is especially true for young children (see Lorch, Bellack, & Augsbach, 1987). In addition, children's ability to successfully produce words tend to vary with an overall verbal ability. Because verbal ability was not measured in the current studies, children with poor verbal ability could have negatively influenced the overall naming score without influencing the overall pointing score. The production data is important because educational videos that aim to build children’s vocabularies are designed to do so by improving both comprehension and production of new words, but it is comprehension that tends to receive the most attention (Sell, Ray, & Lovelace, 1995).
diminished role for the social cues that a speaker commonly provides. However, this difference did not appear to affect children’s ability to quickly learn words. Indeed, the rate of word learning on those trials was similar to the rate found on trials in previous studies that have involved direct interactions between the child and a speaker (i.e., Behrend et al., 2001, Study 1; Carey & Bartlett, 1978). As such, a future question may be how important the presence of a speaker and the presence of the social or pragmatic cues provided by a speaker are if their absence has no noticeable effect on word learning (see Scofield, 2005; Akhtar, 2005 for discussion). At the very least, children’s ability to successfully learn new words in the current studies with diminished social support should be considered impressive.

Also impressive in the current studies was children’s ability to transfer their knowledge of a newly learned word from video to real life, something that is not usually required in typical examples of word learning but is nonetheless important for constructing a lexicon when new words are initially presented via television or video. The complexities inherent in making such a transfer are noteworthy. Children must: (a) apply a word learned for a two-dimensional image to a three-dimensional object (see Schmidt et al., 2007; Zack et al., 2008), (b) tolerate small differences between the image and object (e.g., orientation, shading, etc.), (c) ignore an attraction toward real objects that can be picked up and manipulated, and (d) overcome the difference between the mode of learning and the mode of testing (i.e., children learned a word from video but were asked to point or name an object in real life). These reasons may help account for why previous studies have consistently found that young children, especially children under the age of 3 years old, display the video deficit effect (Deocampo & Hudson, 2005; Hayne, Herbert, & Simcock, 2003; Kirkorian, Wartella, & Anderson, 2008; Troseth & DeLoache, 1998).
This makes 2-year-old’s success in the current studies all the more remarkable. That the video deficit effect was not obviously apparent may be related to the nature of the videos that were used. The current studies used videos with images of objects rather than using segments from educational programming or clips with human actors (e.g., Krcmar et al., 2007; Robb et al., 2009). In addition, the real life objects were exact counterparts to the images seen on the video. Consequently, the images often preserved the size of the objects and always preserved the object’s color and shape (a critical feature when extending words) (Landau, Smith, & Jones, 1988). Finally, it is possible that there are simply some types of information that are less affected by the video deficit effect and words may be one example. Learning words (especially learning nouns) is a common experience for 2-year-olds, one that may be so common as to persist even when involving atypical sources like videos. In the end, using simple videos with a common learning experience may have helped 2-year-olds in the current studies counteract the video deficit effect.

Interestingly, 2-year-olds were also able to transfer their knowledge of a newly disambiguated word from video to real life. Previous studies have shown that when young children learn new words from direct interactions with a speaker, they typically make some basic assumptions about how those words are used, including assuming that words are mutually exclusive. However, it was unclear from these previous studies whether children would honor the same basic assumptions for words learned from video and, in addition, whether such assumptions would survive a transfer real life objects. The results of Study 2 indicated that children were capable of both. Indeed, children in the current studies were able to disambiguate new words at rates that were very similar to the rates found in direct interactions with speakers.

This same reasoning may also help explain why video imitation studies often fail to find the video deficit effect (Hayne et al., 2005).
(Scofield and Behrend, 2007) and there was no significant decline in performance when transferring from video to real life.

However, of some concern is that children’s performance in the disambiguation condition of Study 2 did not replicate children’s performance in a similar disambiguation condition in Scofield and Williams (2009). Scofield and Williams found that children did not successfully disambiguate words learned from video. However, unlike the results of Scofield and Williams, where children disambiguated on only 44% of trials (a percentage that did not differ from chance), Study 2 found that children disambiguated on 76% of trials. To explain their results Scofield and Williams suggested that the small number of trials or the absence of a speaker and speaker-provided cues might explain children’s inability to solve the disambiguation task. The results of Study 2 support the former explanation. Study 2 included more trials than Scofield and Williams (i.e., 4 trials as compared to 2 trials). In addition, Study 2 used a larger sample size (i.e., 32 two-year-olds as compared to 18 two-year-olds), slightly older 2-year-olds (i.e., average age of 33 months as compared to 31 months), and included naming trials which gave children either one or two additional exposures to the disambiguated objects. Of these differences, it is the difference in the number of trials and the inclusion of naming trials that might best explain the dissimilar results between the two studies.

Scofield and Williams (2009) suggested that two trials might be insufficient to bring about disambiguation - especially in the absence of social or pragmatic cues. In support of this suggestion, it is typical for studies that use the disambiguation task to include more than two trials (e.g., Diesendruck & Markson, 2001; Scofield & Behrend, 2007). As such, it is unclear whether the disambiguation pattern revealed after two trials would be the same as the disambiguation pattern revealed after more than two trials. Considering that the logic of the
disambiguation task is complex, especially for young children (i.e., children must learn the word for one object and then infer that a new word refers to a different object despite having never heard a speaker label that new object), disambiguation may be most apparent only on later trials. The results of Study 2 would support this general line of reasoning. However, it is important to note that lexical and social or pragmatic cues are generally considered more theoretically relevant to solving the disambiguation task than is the number of trials (Behrend, 1990; Diesendruck & Markson, 2001; Markman, 1990).

Another possible explanation for the differences between Study 2 and Scofield and Williams (2009) comes from including real trials (i.e., trials during which children selected real objects). Including real trials might have inflated children’s performance on the disambiguation task since children’s selection on both the real trials and the video trials counted toward their overall performance. This essentially gives children an extra chance to solve the task and an additional exposure to the word and objects. In contrast, Scofield and Williams did not use real trials and so only video trials counted toward children’s overall performance. In addition, the real trials in Study 2 required the experimenter to interact with the children thereby introducing a social element to Study 2 that was not present in Scofield and Williams. Each of these possibilities, either considered separately or considered together, could help explain the dissimilar results across the two studies. Ultimately however, the safest conclusion is that the results of the two studies are inconsistent and do not sufficiently resolve the question of whether children can disambiguate words learned from video.
CHAPTER 5

CONCLUSIONS

The current studies set out to determine whether children could successfully learn words from video (after limited exposure), add those words to their productive vocabularies, transfer the word from video to real life, and disambiguate a newly learned word. The results showed that children were able to learn and produce words from video, transfer that knowledge to real life objects, and disambiguate the newly learned word. It is important to note that these studies looked only at children’s ability to learn words for objects and that these results may not transfer to other types of words (e.g. verbs and adjectives) or other types of information (e.g., facts). That said though, these results are impressive given the diminished social cues that normally assist children when learning words (Scofield et al., 2007; Scofield & Williams, 2009) and also because children in this age range often exhibit the video deficit effect, a failure to transfer information from a video to real life. Although more research is certainly needed, the results of the current studies suggest that videos can be an effective source for learning words.
REFERENCES


Table 1

*Pointing and Naming in Study 1 and Study 2 (out of 4 Trials)*

<table>
<thead>
<tr>
<th>Condition</th>
<th>Study 1</th>
<th>Study 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Video Pointing</td>
<td>2.38 (60%)</td>
<td>3.03 (76%)</td>
</tr>
<tr>
<td>Video Naming</td>
<td>1.72 (43%)</td>
<td>1.94 (49%)</td>
</tr>
<tr>
<td>Real Pointing</td>
<td>2.06 (52%) a</td>
<td>2.81 (70%) a</td>
</tr>
<tr>
<td>Real Naming</td>
<td>1.84 (46%)</td>
<td>2.22 (56%)</td>
</tr>
</tbody>
</table>

*a different from chance, $p < .05$
APPENDIX B

FIGURES
**Figure 1.** Pointing and Naming in Study 1.

<table>
<thead>
<tr>
<th>Familiarization</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Image" /></td>
<td>“This is a fep! Look at the fep! It’s a fep!”</td>
</tr>
<tr>
<td>Test Trials</td>
<td>Video Pointing</td>
</tr>
<tr>
<td><img src="image2" alt="Image" /> <img src="image3" alt="Image" /> <img src="image4" alt="Image" /> <img src="image5" alt="Image" /></td>
<td>“Which is the fep? Can you point to the fep? Point to the fep.”</td>
</tr>
</tbody>
</table>

**Figure 2.** Pointing and Naming in Study 2.

<table>
<thead>
<tr>
<th>Familiarization</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image6" alt="Image" /></td>
<td>“This is a vox! Look at the vox! It’s a vox!”</td>
</tr>
<tr>
<td>Test Trials</td>
<td>Video Pointing</td>
</tr>
<tr>
<td><img src="image7" alt="Image" /> <img src="image8" alt="Image" /></td>
<td>“Which is the hap? Can you point to the hap? Point to the hap.”</td>
</tr>
</tbody>
</table>